

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2001, 23:47:01 ; Search time 6769.47 Seconds
(without alignments)
4526.442 Million cell updates/sec

Title: US-09-075-375A-1
Perfect score: 1981
Sequence: 1 tgggggttcgaattttacc.....aattttggcctaaaaaaa 1981

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vil:*
59: gb_vil2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rol:*
95: gb_rod:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1759	88.8	1760	14	LSU31462	U31462 Lactuca sat
2	643.8	32.5	1589	14	NTU34817	U34817 Nicotiana t
3	621	31.3	1686	15	SOL250433	AJ250433 Spinacia
4	611.8	30.9	1555	14	ATU44133	U44133 Arabidopsis
5	413.2	20.9	1173	13	AF288196	AF288196 Oryza sat
6	406.4	20.5	133843	12	AC003981	AC003981 Genomic s
c 7	404.8	20.4	2884	14	ATU39452	U39452 Arabidopsis
8	143.2	7.2	89479	12	AC006932	AC006932 Genomic s

```

c 9 78 3.9 7218 10 I66494
c 10 60.4 3.0 1141 10 AX083744
c 11 56.2 2.8 4788 94 AY008297
c 12 56.2 2.8 201978 76 AC079533
c 13 55.4 2.8 156610 75 AC074378
c 14 55 2.8 39987 4 AC009601
c 15 55 2.8 42160 4 AC005804
c 16 53.8 2.7 204099 69 AC025669
c 17 53.4 2.7 973 53 CNS071LE
c 18 53.2 2.7 204847 87 AC018664
c 19 52.8 2.7 1141 10 AX083744
c 20 52.6 2.7 108409 58 AF083424
c 21 52.2 2.6 214194 75 AC073757
c 22 52 2.6 147243 84 HSAJ3612
c 23 51.8 2.6 118593 92 HS288L1
c 24 51.6 2.6 223538 77 AC087150
c 25 51.4 2.6 169931 86 AC005822
c 26 51 2.6 200345 68 AC023427
c 27 50.6 2.6 15974 87 AC018647
c 28 50.4 2.5 1228 7 MRU59897
c 29 50.4 2.5 166735 79 AL355861
c 30 50.2 2.5 196998 74 AC069113
c 31 50.2 2.5 215478 78 AF276759
c 32 50.2 2.5 215963 71 AC040982
c 33 50 2.5 197886 67 AC022779
c 34 49.8 2.5 200345 68 AC023427
c 35 49.6 2.5 157921 71 AC027726
c 36 49.6 2.5 166871 61 AC010647
c 37 49.6 2.5 169212 81 AL390958
c 38 49.6 2.5 185931 76 AC083812
c 39 49.4 2.5 13095 93 HSFGR4G
c 40 49.4 2.5 218846 81 AL589701
c 41 49.2 2.5 105567 91 AP001069
c 42 49.2 2.5 167854 80 AL358235
c 43 49.2 2.5 169118 79 AL162502
c 44 49.2 2.5 192094 73 AC068640
c 45 49.2 2.5 245802 60 AC006279

```

ALIGNMENTS

```

RESULT 1
LSU31462 1760 bp mRNA PLN 22-JUL-1996
LOCUS Lactuca sativa violaxanthin de-epoxidase (VDEL) precursor, mRNA,
DEFINITION complete cds.
VERSION U31462
KEYWORDS U31462.1 GI:1438874
SOURCE Lactuca sativa.
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
II; Asterales; Asteraceae; Lactuceae; Lactuca.
REFERENCE 1 (bases 1 to 1760)
Bugs, R.C. and Yamamoto, H.Y.
Molecular cloning of violaxanthin de-epoxidase from romaine lettuce
and expression in Escherichia coli
Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6320-6325 (1996)
96270536
REFERENCE 2 (bases 1 to 1760)
Bugs, R.C.
Direct Submission
Submitted (12-JUL-1995) Plant Molecular Physiology, University of
Hawaii, 3190 Maile Way, Honolulu, HI 96822, USA
Location/Qualifiers
1. .1760
/organism="Lactuca sativa"
/strain="Romaine"
/db_xref="taxon:4236"
transit_peptide 235..609
/gene="VDEL"

```

```

CDS
235..1656
/gene="VDEL"
/codon_start=1
/product="violaxanthin de-epoxidase precursor"
/protein_id="AAC49373.1"
/db_xref="GI:1438875"
/translation="MALSLHTVFLCKEALNLYARSPCNERFHRSGQPTNIMMKIR
SNNGYNSFRLFTSYKTSFSDSHCKDKSKQICSDTSPETQRFDLKRGMTLLLEKQ
WFOFOLAIVLVCTFVIVPRVDALAKTACLLKCEKTELAKCIANPSCAAVACLO
TCNNRDETECQIKCCDLEFNSVVDQFNCAVSRKCKVPRKSDVGEFPVDRNAVQV
FNKKPDSGKWTITSLNPTFDADFCDLHEFHENDKLVNLRKTLDDGGFTSAY
OTFQVQDPLPGALYNHNDNEFLHYQDDWYLLSIOENKLPDDYLFVYRGRNDADWYGG
SVIYTRSPILPESIIIPNLQAAKSGVRDNFNTITDNSCGPEPPLVERLEKTAEGEK
LLIKEAVEIEEEVEKEVRDTEMTLFQRLLEGFKELQDQDEENFVRELSEKEEKLIN
ELQMEATEVEKLFGRALPIRKLR"
235..1656
/gene="VDEL"
mat_peptide
610..1653
/gene="VDEL"
/product="violaxanthin de-epoxidase"
BASE COUNT 557 a 319 c 387 g 497 t
ORIGIN
Query Match 88.8%; Score 1759; DB 14; Length 1760;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 tctgggttcgaattttaccaccacacaaagtgttgcctaccataattgggataagagctct 60
Dd 1 TctGGGtTCGAAATTTTACCACACCAAGATTTTGTCTTACCATAATTGGGATAAGAGTCT 60
QY 61 aatttccctgtacaaattttccaaatttctctcctccgcacacacatatatactgtacgc 120
Dd 61 AATTTCCTTGTACAAATTTTCCAAATTTCTCTCCGCGCACACCATATATACTGTACGC 120
QY 121 cacttcgaacgtcacaattgttgaaaaaagacagcagattttacaaacacgagaagataa 180
Dd 121 CACTTCGAACGCTACAATGTTTGAAAAAAGACGAGATTTTACAAAGACGGAGAGATAA 180
QY 181 laagcttcaagtactccgactcgcaggtggcctttggaagcaacaaactggctatggct 240
Dd 181 TAAGCTTCAAGTACTCCGATCGTCAGGTGGCTTTTGGAAAGCAACAACTGGCTATGGCT 240
QY 241 ctttcttcacactgtattctctgcgaagagaaagccctcaatttatatgcagaatca 300
Dd 241 CTTTCTCTCACACTGTATTCTCTGCAAGAGAGAAAGCCCTCAATTTATATGCAAGATCA 300
QY 301 ccattgaatgaaagggttccacaggagtgcacaaacctcctaccacacataatcatgatgaa 360
Dd 301 CCATGTAATGAAGGTTTCACAGGAGTGCACAACTCTCTACCAACATAATCATGATGAAA 360
QY 361 attgatccacaacatggatatttaattcttccgttgggtttgatttatataacacaaagt 420
Dd 361 ATTCGATCCAAACAATGGATATTTTAAATTTCTTCCGGTGTGTTTACATCTTATAAGCAAGT 420
QY 421 tcttctcagattctagcattgcgaagataaattctcagatatgcagcatgcagacaaagt 480
Dd 421 TCTTCTCAGATTCTAGCCATTGCAAGGATRAATCTCAGATATGCAGCATGCATACAAGT 480
QY 481 tttaggagaaatacaagaatttgatctcaaaaggggcagtcgactttgattcttgaaaaagcaa 540
Dd 481 TTTAGGAGAAATACAAAGATTTTGATCTCAAAAGGGGCATGACTTTTGAATCTTTGAAAAGCAA 540
QY 541 tagagacaattcataaattggctatcgtattgttgcacatttggttatcgttccacaga 600
Dd 541 TGGAGACAAATTCATCAATTTGGCTATCGTATTTGGTTTGCACATTTGTTATCGTCCACAGA 600
QY 601 gttgatgccgtttgatcgtctttaaacttgcctgtgttttactcaagaagatgcagattgag 660
Dd 601 GTTGATGCGGTTGATGCTCTTAAAAACTTGTGCTTCTTTACTCAAAAGATGCAAGGATTGAG 660
QY 661 cttgcaaaatgatagcaaaacccctctgtgcggcgcaaacgttgcctgtctacagactgc 720

```

```

Db 661 CTTCAAAATGTATAGCAAAACCCATCTGTGCGCAAAACGTTGCCGTGTACAGACTTC 720
|
Qy 721 acaatcgtcctgacgagaccgaatgtcagataaaatgtgtgactgtgtcgaataacagt 780
|
Db 721 AACAAATCGTCTGACGAGACCGGAATGTACAGATAAAATGTGTGACATGTTTCGAACAGT 780
|
Qy 781 gtgtgagacaattcaacagagtggtggttcccgaaagaaatgtgtgccccggaatcg 840
|
Db 781 GTGTGTGACCAATTCACAGAGTGTGCGGTTCGCCGAAGAAATGTGTGCCCGGAAATCG 840
|
Qy 841 gatgtgtggaattcccggttccggtcgtatcgtatcgtgtgtcgaataatttaacatgaa 900
|
Db 841 GATGTGGGTGAATTCGCGGTTCGGATCGTAATGCAAGTGTGTCAAAATTTTAAACATGAA 900
|
Qy 901 gactttagtgtggaagtgttatatacaaatgtgtttaaactcctacattgtgacattgat 960
|
Db 901 GACTTTAGTGGGAAGTGTATATAACAAAGTGTGTAAATCCTACATTTGATGCAATTTGAT 960
|
Qy 961 tgtcaactcatgagttcatatggaataatgataaaactgtgtgggaacttaacatgagcgc 1020
|
Db 961 TGCAACTTCATGAGTTTCATATGGAATATGAAATGATAAACTTTGTTGGGAACCTTAACATGCCGC 1020
|
Qy 1021 ataaaacttggatggtgttcttactcgtatcgtatcgtgtgcaaacatttctcaaat 1080
|
Db 1021 ATAAAACCTTTGGATGTTCTTTACTCGATCTGCTGTGCAAAACATTTGTTCAAGAT 1080
|
Qy 1081 ccagattctcgtgagcaactttataatcatgacaatgagtttcttaccacaaagatgac 1140
|
Db 1081 CCAGATCTTCTGGAGCACTTTATAATCATGACAAATGAGTTTCTTCACTACCAAGATGAC 1140
|
Qy 1141 tggatcatattcttcccaatcgaataacccgcatgattacataattctgtactac 1200
|
Db 1141 TGATACATATTCTTCCCAATTCGAAACAAACCCGATGATTACATATTCCTATACTAC 1200
|
Qy 1201 cgaggtcgaaacgacgcatggatgatacgtgtggtggtcgtatctacacccgaagccgc 1260
|
Db 1201 CGAGGTTCGAAACGACGATGGATGATACGTTGGTCCGTCATCTACACCCGAAAGCCGC 1260
|
Qy 1261 acactcccgaaatcgatcaccacacccacaaagcgaacaaatcgtggtcgagac 1320
|
Db 1261 ACATCTCCCGAAATCGATATCTCCAAACCTTACAAAGACGACCAATTCCTGTCGAGAC 1320
|
Qy 1321 tttaacaatttcataaaccacacgaatagtgtggcctgagcctccattgtgaaagg 1380
|
Db 1321 TTTAACAAATTTTATACAAACCGACATAGTTTGGCCCTGACCCCTCCATTTGTTGGAAAGG 1380
|
Qy 1381 cttgagaaacacgaggaagggcgagaagttgttgataaaagaagctgtagagatagaa 1440
|
Db 1381 CTTGAGAAACACGCGGAGAGGCGGAGAAAGTTGTTGATAAAAGAAAGCTGTAGAGATAGAA 1440
|
Qy 1441 gaagaggttgaaaaagagtgagaaggttagagatactgagatgacitgtttcagagg 1500
|
Db 1441 GAAGAGTTGAAAAAGAGTGGAGAGGTTAGAGATAGTGTGATGACTTTGTTTCAGAGG 1500
|
Qy 1501 ttgcttgaaggggttgaagagttgcaacagatgaagagaatttggtagggaggtgagt 1560
|
Db 1501 TTGCTTTGAAGGGTTTAAAGGAGTTTGAACAAGATGAAGAGAATTTTGTGAGGAGTTGAGT 1560
|
Qy 1561 aaagaagaagaagaattctgaatgaacttcaatggaagcgaactggaagtgaaagcgtt 1620
|
Db 1561 AAAGAAGAAGAAGAAATTTCTGAATGAACCTTCAATGGAAGCGACCTGAAGTTGAAAGGCTT 1620
|
Qy 1621 ttgtggcgcgctgtaccgattggaacttagataaaatttgatgattgattcagacaaat 1680
|
Db 1621 TTTGGCGCGCGTGTACCGATTAGGAACCTTAGATAAAATTTTCGATGATTGATTCAGACAAAT 1680
|
Qy 1681 atatatgtcatatggtattatagatactagagaaaccccaaaaacttttggatatagc 1740
|
Db 1681 ATATATAGTCATATGATATGATAGATAGTATGATAGAGAAACCCCAAAAACCTTTTGTATACG 1740
|
Qy 1741 tgataaacgtgtttgtgat 1759
|

```

```

Db 1741 TGATAAACGTGTTTGTGTAT 1759

RESULT 2
LOCUS NTU34817 1589 bp mRNA PLN 15-JUN-1998
DEFINITION Nicotiana tabacum violaxanthin de-epoxidase precursor (TVDE1),
mRNA, complete cds.
ACCESSION U34817
VERSION U34817.1 GI:1463122
KEYWORDS common tobacco.
SOURCE Nicotiana tabacum
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Nicotiana.
REFERENCE 1 (bases 1 to 1589)
AUTHORS Bugos,R.C.; Hieber,A.D. and Yamamoto,H.Y.
TITLE Xanthophyll cycle enzymes are members of the lipocalin family, the
first identified from plants
J. Biol. Chem. 273 (25), 15321-15324 (1998)
98288256
REFERENCE 2 (bases 1 to 1589)
AUTHORS Bugos,R.C. and Yamamoto,H.Y.
TITLE Direct Submission
SUBMITTED (25-AUG-1995) Robert C. Bugos, Plant Molecular
Physiology, University of Hawaii, 3190 Maile Way, Honolulu, HI
96822, USA
FEATURES
Location/Qualifiers
source 1..1589
/organism="Nicotiana tabacum"
/strain="Xanthi"
/db_xref="taxon:4097"
/tissue_type="leaf"
gene 1..1589
/gene="TVDE1"
CDS 42..1478
/gene="TVDE1"
/codon_start=1
/product="violaxanthin de-epoxidase precursor"
/protein_id="AAC50031.1"
/db_xref="GI:1463123"
/translat="MALAPHSNFLANHETIKYVYKSLPGHKRFSNGWEDYFGSIVVA
KICSSRRIPRYFRKSPRICGLDSRGLQFVSHKHNLSPAHSINQVPGNSCKFPK
DVALMWKRWGQFAKTAIVAFILSVASKADAVDALKTCTLLKECLLELAKISNPA
CAANVACLOQFNRRPDETECIKGLFENSVDEFNECAVSRKCVPRKSDVGDPP
PDSVLVQKFDNMDFSGKWFITRLNPTDFADFCQLEHTEENKLVGNLSWRITPD
GGFTRSVAVOKFVQDPKYPGILYNHONEYLLYODDWY ILSSKVENSPEDYIFVYKGR
NDAMDYGGSVLYTRSAVLPEIIPLOTAOKVGRDENTFTKTDNTCGPEPLVRL
EKVVEGERTIIKEVEIEEVEKVRDKEVTLFSLFEGFKELQRODENFLRELSKEE
MDVLDGLKMEATEVEKLFGRALPIRLR"
transit_peptide 42..443
/mat_peptide /gene="TVDE1"
444..1475
/gene="TVDE1"
/product="violaxanthin de-epoxidase"
BASE COUNT 500 a 262 c 366 g 461 t
ORIGIN

Query Match 32.5%; Score 643.8; DB 14; Length 1589;
Best Local Similarity 71.88; Pred. No. 2.5e-126;
Matches 862; Conservative 0; Mismatches 327; Indels 12; Gaps 1;

Qy 483 tgaggaatacaagaattgatctcaaaaggcgcatgatttatttatttgaagaagaatg 542
|
Db 317 TAAGGGAATTCAGGATGCAAAATTCACAAAGATGTAGCTTTGATGCTTTGGGGAATG 376
|
Qy 543 gagacaattcatacaattggctatcgattggttgcacattgttatcgatctccagagt 602
|
Db 377 GGGCAATTTGCCAAAACAGCAATTTAGTATATTCATTTTGTCTGCTTCAAAAGC 436
|
Qy 603 tgaatgcgtgtgatgctcttcttaaaactgtgttattactcaagaagaatgcaggattgagct 662
|

```

```

Db 437 TGAATGGGTGATGCTCTCAAGACTTGTACTTCTTACTGAAGAGTGCAGGTAGAGCT 496
QY 663 tgcaaatgtatagcaaaacccatctgtgtgagcaaacgttgcctgtctacagacttgaa 722
Db 497 TGGGAAGTGATTTGCAACCCCTCATGTGAGCTAATGTGCTGTCTCAGACTTGCAA 556
QY 723 caatgctctgacagacgaatgtcagataaaatgtgtgactgtgtcgaacacagtg 782
Db 557 CAATAGACCTTGACGAACGAATGTGAGATAAATGTGTGATTTGTTTCAAAACAGTGT 616
QY 783 ggtggaccaattcaacgagtgctggtgttcccgaaagaaatgtgtgccccggaacatcga 842
Db 617 CGTAGACGAGTTCATGAGTGTGACGTCTCCCGAAAGAAATGTGTACCTCGTAATCTGA 676
QY 843 tgtgggtgaattcccggttcccgatgtaacgagtggttccaaattttaacatgaaga 902
Db 677 TGTGTGTGACTTCTCTGCTACCTGCTCCAGTGTCTTCTCCAGAGTGTGACATGAAGA 736
QY 903 cttagtggaagtgtatatacaagtggtgtttaaactctacattgtgacattgattg 962
Db 737 TTTTAGCGGAATGTTCTATTAATCGCGGTTCGATCCACTTTTGATGCTTTTGATTG 796
QY 963 tcaacttcattgatttcataatggaatgataaaactgtgtggaaacttaacatggcgat 1022
Db 797 CCAATTGCATGAGTTCATACAGAAGAAACAAACTTGTGGGAATTTATCTTGGAGAAT 856
QY 1023 aaaaacttgatggtgtgttcttactgactgtgtgcaaacattgttcaagatcc 1082
Db 857 ACGTAGACCTGTGAGGAGATTTTACTCGATGAGCGGTGCAAAATTCGTGCAAGATCC 916
QY 1083 agatctccctggagcatttataatcatgacaatgagttcttccactaccagagactg 1142
Db 917 AAGTATCCGGGATACTCTACATCATGATATGAGTATCTTCTACCAAGTACACTG 976
QY 1143 gtacatatattcttcccaatcgaaacaaacccgatgattacatatcgtatactaccg 1202
Db 977 GTATATTTTGTATCCCAAGTAGAAAATAGTCCAGAGGATTACATATTTGTGTACTATA 1036
QY 1203 aggtcgaacagcagcatgtagatgacgtggtgcgtgactacacccgaagcccgac 1262
Db 1037 GGGCAGAAATGATGCTGGGATGATGATGTTGTTCTGTACTTACTACAGAAGTSCAGT 1096
QY 1263 actcccgaaatcgatcatcccaacactacaaaagcagcaaatccgtgggtcgagact 1322
Db 1097 TTTGCTGAAAGCATTTATACGGAGTTGCAACCGCAGCTCAAAAAGTTGGCGTGTATT 1156
QY 1323 taacaatttcataaaccagcaaatagttgtggcctgagcctccatgtgtgaaaggct 1382
Db 1157 CAACACATTTATAAACAACAGACAATACATGTGGCCCTGAAACCTCCCTTTGAGAGGTT 1216
QY 1383 tgagaaacagcggagagggcgagaagttgttgataaaagaagctgtagagataaga 1442
Db 1217 GGAGAGAAAGTGAAGAGAGAGAGAGAGCATATAAAGAGATTGAGGAGATAGAAGA 1276
QY 1443 agaggttgaaaaagagtgagagaggttagagatactgagatgactgtgttcagaggtt 1502
Db 1277 AGAAGTA-----GAGAGAGGTGAGAGATGAAGAGAAAGTCAACCTTATTCAGTAACT 1324
QY 1503 gcttgaagggtttaagaggttgcaacaagatgaagaaatttggaggaggttgagtaa 1562
Db 1325 GTTTGAGAGGTTTTAAGAGCTCCCAACGAGATGAAGAGAACTTCTTAAGAGAGGTGAGCAA 1384
QY 1563 agaagaaagaaattctgaaatgaacttcaaatggaagcagactgaagttgaaagctttt 1622
Db 1385 AGAAGAAATGATGTTTGGATGACCTTAAATGGAAGCAACTGAGGTAGAAAAACTTTT 1444
QY 1623 tggcgcggttaccgattaggaacttagataaaatttcagattgattcagacaatat 1682
Db 1445 TGGCGGTGCTTTTACCAATAAGGAATTAAGGTAAGTATTTTAAACTATCAACATATAT 1504
QY 1683 a 1683

```

```

Db 1505 A 1505

RESULT 3
LOCUS SOL250433 1686 bp mRNA PLN 21-OCT-1999
DEFINITION Spinacia oleracea mRNA for violaxanthin de-epoxidase (svdel gene).
ACCESSION AJ250433
VERSION AJ250433.1 GI:6103242
KEYWORDS svdel gene; violaxanthin de-epoxidase.
SOURCE Spinach.
ORGANISM Spinacia oleracea
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales;
Caryophyllales; Chenopodiaceae; Spinacia.
REFERENCE 1 (bases 1 to 1686)
AUTHORS Emanuelsson, A.K., Eskling, M. and Akerlund, H.E.
TITLE Cloning and sequencing of Spinacia oleracea violaxanthin
de-epoxidase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1686)
AUTHORS Emanuelsson, A.K.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1999) Emanuelsson A.K., Dept. of Plant
Biochemistry, Lund University, Chemistry Center, POB 117, 221 00
Lund, SWEDEN

FEATURES
source 1..1686
/organism="Spinacia oleracea"
/db_xref="taxon:3562"
/country="Sweden"
/tissue_type="leaf"
sig_peptide 1..372
/gene="svdel"
CDS 1..1419
/gene="svdel"
/function="part of the xanthophyll cycle"
/codon_start=1
/product="violaxanthin de-epoxidase"
/protein_id="CAB59211.1"
/db_xref="GI:6103243"
/translacion="MALVARSCVSYVEIAGICNNVSHRNFKKVVQWKNPFLPQDDAR
RNIFNDRKLSCTKFCIGASEKLOHSPKSGLLSCGWEVNSKVVSNVPIPKWNLK
LKVEVTAIVACTFFVYSSQAQVDALCTCLLEKRIELAKCIANPSCAANVACLOT
CNNPDETQICGDLDFANQVDFNECAVSRKKCPQKSDGVEPDPSPVLYKSF
NMADFNKGWFISSGLNPTDFAPQQLHEFDGLKVLGNLSWKIKPTDGGFFTRAVQ
KFAQDPSPQGLYVHNHAYLHQQDYILSKTENQDDYFVYVYRGRNDWDGYYGA
FLYRSATVPENIVPELNRAAQSVGKDFNFKTIDNTCGPEPLVERLEKTEVEGERT
IIKEVEOLEGEIEGDLKVGKTEMTLQRLLEGFQELQKDEEYFLKELNKEERELLED
LKMDAGEVKLFGLALPIRLR"
gene 1..1419
/gene="svdel"
mat_peptide 373..1416
/gene="svdel"
/product="violaxanthin de-epoxidase"

BASE COUNT 549 a 296 c 375 g 466 t
ORIGIN

Query Match 31.3%; Score 621; DB 15; Length 1686;
Best Local Similarity 72.5%; Pred. No. 1.7e-121;
Matches 804; Conservative 0; Mismatches 305; Indels 0; Gaps 0;

QY 557 aattggctatcgattggtttgcacattgttatcgttcccgagagttgattcattgag 616
Db 320 AAGTGACTGCTATAGTTCGGTGTACATTTTGTTCATGCTTCGGCACAAGCTGTTGATG 379
QY 617 ctcttaaaactgtctgttcttactcaagaatgagcaggttgagcttgcaaaatgtatg 676
Db 380 CTTCTAAACTTGTACCTGTTGCTGAGGAATGCAGAATTGAACCTGCCAAGTGCATCG 439
QY 677 caaacccatctgtgctgagcaaacgttgcctgtctacagacttgcaaacatgctcagc 736

```


Db	440	CAAAATCCATCTGTGCGAGCCAAATGTTGCTTGCGCTGCAGACCTCGCAATATAGACCTGATG	499
QY	737	agacggaatgcagataaaatgtgtgactgtgtcgaacacagtggtggtggaaccaattca	796
Db	500	AAACCGAATGCCAAATCAAATGTTGGGACTGTGTCGCCAACAAAGTGTGTGACGAGTTCA	559
QY	797	acgaagtgcgggtttcccgaaagaaatgtgtcccccggaaatcggatgtgggtgaattccc	856
Db	560	ACGAGTGTGCAGTCTCTCGAAAGAAATGGTGCTCAGAAGTCTGATGTTGGAGAATTTTC	619
QY	857	cggttccggatcgtcaatgcagtggtttcaaaattttaacagaaagacittagtggggaagt	916
Db	620	CTGTTCCCGATCTTAGTGTGCTCGTTAAGAGTGTCAACATGGCAGATTTTCAACCGGGAAGT	679
QY	917	ggatatacaagtggtttaaatccacatttgatgcatgttgatgtcaacttcgatagtt	976
Db	680	GGTTTATAAGTAGTGTGCTAAACCCCTACATTCGAGCGCTTTTGATTGCCAGTTTACATGAGT	739
QY	977	ttcatatggaaaatgataaacttgttggaaacttaacatggcgcataaaactttggagt	1036
Db	740	TCCATTTTGGGAAGATGGAAAACTTTGTTGGAAACTTGTCTTGGCGAATAAAACACCGATG	799
QY	1037	gtggtttcttactcgatcgtgtgtcgaacatttgttcaagatccagatccttccatgag	1096
Db	800	GTGGTTTTTTCACACGCACTGCTGTACAGAAATTTGCGCAAGACCCCTCTCAACCTGGAA	859
QY	1097	cactttaatcatgacaatgagttcttccactccaagatgactggtcacatatatctt	1156
Db	860	TGCTGTATAATCATGACAATGCATATCTTCACTATCAGATGATGTGTACATCTCTATCTT	919
QY	1157	ccaaatcgaacaaacccgatgtattcatattcgtatactaccgaggtgcgaacgacg	1216
Db	920	CTAAATTTGAAATCAACCAGATGACTAGTATTTTGTATATTACCGAGGCAGGAATCATG	979
QY	1217	catgggatgatcagtggtgcggtgatctacacccgaagcccgacactcccgaaatcga	1276
Db	980	CGTGGNATGCTATGTGTGGCGATTTCCTTTACACAAAGAAAGTCCAACCTGTACTTGAAATA	1039
QY	1277	tcatcccaaacctcacaaaagcagcgaatccgtgggtcggagactttaaceatttcaata	1336
Db	1040	TCGTGCTGAACCTTAACAGACGAGCTCAAGGTGTAGGAAAAGACTTTCANATTAATTCATCC	1099
QY	1337	caaccgacaatagttgtggccctgagctccattggttgaaaggtcttgaaacacgacgg	1396
Db	1100	GAACGGACAATACCTGTGGGCCGAGCTCCACTGGTTGAGAGGCTGGAGAAGCGGTGG	1159
QY	1397	aagaggcgagaaagtgttgataaaaagcgtgtagatagaagaagaggttgaaaaag	1456
Db	1160	AGGAAGSAGAGAACAACATCATTTAGGAAGTGAACAATTTAGAAGGAGAGATAGAAGGAG	1219
QY	1457	agtggaaggttagagatactgagatgactttgtttccagaggttgcgttgaaaggttta	1516
Db	1220	ACCTGGAAAAGGTTGGGAAAACGTAGATGACATTTGTTTCAGAGGCTACTAGAGGTTTTC	1279
QY	1517	agaggtgcaacaagatgaagaaattttgtgaaggagttgtgaaggagtttgaaaggaataa	1576
Db	1280	AAGAGCTTCAAAGATGAAGAATACTTCTCTCAAGAAATTTAAACAGAGGAAGACGAGAAT	1339
QY	1577	ttctgaatgaacttcaaatgggaagcgactgaagtgtgaaagcctttttggcgcgcttac	1636
Db	1340	TACTGGAACACTTAAAGATGGACGCTGGCGAGGTTGGAAAAACTATTTTGGCTAGCATTTAC	1399
QY	1637	cgattaggaaacttagataaaatttcgatg	1665
Db	1400	CTATAAGAAGCTTCGGTAAACATTTTCAGG	1428

RESULT	4
ATU44133	
LOCUS	ATU44133 1555 bp mRNA PLN
DEFINITION	Arbidopsis thaliana violaxanthin de-epoxidase precursor (AVDE1)
ACCESSION	mRNA, complete cds. U44133

U44133.1 GI:1465734
 .
 thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsi
 1 (bases 1 to 1555)
 Xanthophyll cycle enzymes are members of the lipocalin family, the
 first identified from plants
 J. Biol. Chem. 273 (25), 15321-15324 (1998)
 98288256
 2 (bases 1 to 1555)
 Bugos,R.C. and Yamamoto,H.Y.
 Direct Submission
 Submitted (02-JAN-1996) Robert C. Bugos, Plant Molecular
 Physiology, University of Hawaii, 3190 Maile Way, Honolulu, HI
 96822, USA

FEATURES
 source Location/Qualifiers
 1..1555
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 gene 1..1555
 /gene="AVDE1"
 CDS 45..1433
 /gene="AVDE1"
 /codon_start=1
 /product="violaxanthin de-epoxidase precursor"
 /protein_id="AAC50032.1"
 /db_xref="GI:1465735"
 /translation="MAVATHCTSPCHDRIRFFSDDGIGRLGTRKRINGTFLKLKIL
 PPLQSRADRTGRSGRSLPSAFRSFGSKIFDIPVLPSKNELKEITAPILLKLGVLA
 IAGLIPFSAVDADVALTCAACKLGCRIELAKTIANPACAAANVACLQTNNRPDETQC
 IKGLIVSFNVDVEFECAVRKKCVPRKSDLGEFPADPVSIVONFNISDFNGKWYI
 TSLGNPTDFAFCQLHEFTTEGDKNLGVNSWKRIKTLDSGFTTSRAVKVFQDDNPQG
 VLYNHNEYLHQDDMYLLSSKIENKPEDY IFVIYGRNDANDGYGGAVVYTRSSVL
 NSIIPELEKAASIGRDFSFTFRDTONTGCEPALVERIKTVBERGERIIVKEVEIEE
 EVKEVEKVGREMTLPQLRALGCFNELKODEFNFVELSKEEFDELKMEASEVEKE
 LFCKALPIRKVR"

transit_peptide 45..383
 /gene="AVDE1"
 mat_peptide 384..1430
 /gene="AVDE1"
 BASE COUNT 485 a 285 c 376 g 409 t
 ORIGIN

Query Match		30.9%;	Score 611.8;	DB 14;	Length 1555;
Best Local Similarity		72.5%;	Pred. No. 1.5e-119;		
Matches	806;	Conservative	0;	Mismatches 302;	Indels 3; Gaps 1;

```

QY 550 ttcatagaatggtcattgattggtttgcacatttgattgattccagagtgtgatgcc 609
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 324 TTGCTAAAACTCGTGGGTGTTTAGTGTCGCCGTTCCTATTGTTCCATCTCGACATGCA 383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 610 gttgatgtctttaaaactgtgctgttactcaagaatgcaggattgagcttcacaaa 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 GTTGATGCACCTTAAAACTTGTCATCTTATTCAAGGGATGCAGGATAGAATCGCAANG 443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 670 tgtagtaaacaccaccttgtcgggcaaacgctgtcctgtctacagacttgcacaatogt 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 TGCAATTGCCAACACCCTGCCCTGTGCAGCCAATGCGCGTGCCTTCACACCTGCAATAACCGT 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 730 ctgcagcagaccaaatgtcagataaaatgtgacttgttccgaacacagttgtgtgac 789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 504 CCAGATGAACACGAGTCCAGATTAAATGTGGGGATCTGTTTGACAACAGTgttGTTGAT 563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 790 caattcaacagatgtcgggtttccccgaagaatatgtgtcccccgaaatcggatgtgggt 849
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 564 GAGTTCACAGAGTGTCTGTGTCGAAAAAAGTGTGTCTTAGAAAAATCTGATCTCGGA 623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

TITLE		Molecular cloning and expression of rice violaxanthin de-epoxidase	
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1173)		
AUTHORS	Lin,R.-C., Xu,C.-C., Li,L.-B. and Kuang,T.-Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-JUL-2000) Photosynthesis Research Center, Institute of Botany, Chinese Academy of Sciences, Xiangshan, Beijing 100093, China		
REFERENCE	3 (bases 1 to 1173)		
AUTHORS	Lin,R.-C., Xu,C.-C., Li,L.-B. and Kuang,T.-Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (18-SEP-2000) Photosynthesis Research Center, Institute of Botany, Chinese Academy of Sciences, Xiangshan, Beijing 100093, China		
REMARK	Sequence update by submitter		
COMMENT	On Sep 18, 2000 this sequence version replaced gi:9716690.		
FEATURES	Location/Qualifiers		
source	1..1173		
	/organism="Oryza sativa subsp. indica"		
	/cultivar="zhenhui-249"		
	/sub_species="indica"		
	/db_xref="taxon:39946"		
	<1..870		
	/note="RVDE"		
	/codon_start=1		
	/product="violaxanthin de-epoxidase"		
	/protein_id="AAF97601.2"		
	/db_xref="GI:10181228"		
	/translation="DEFNECAVSRRKKYPOKSDYGEFPVPDPVSALVKNFMADPNKGMALYNSHGNPTFDLCOLHEFRVVEGDKLIANTWRITPDSGFFRTALORVPOADPAILYNNHNEFLHYQDDWYIISCKNEKDDYIFVYIRGRNDWADGYGAVLYITRSKYVPELIVPELARAASVGRPFSTIRTDNCGPEPLVERIKTEIKTEVQGEKTIIRVEQVLEIETEGEVKLEEEVEVTLFKRLITDGLMEYKQDLMNPFQGLSKEMEMLLDQNMNEATEVEVFKVFSRALPIRKLR"		
CDS	367 a 255 c 282 g 269 t		
BASE COUNT	367 a 255 c 282 g 269 t		
ORIGIN			
	Query Match 20.9%; Score 413.2; DB 13; Length 1173;		
	Best Local Similarity 66.3%; Pred. No. 1.7e-77;		
	Matches 595; Conservative 0; Mismatches 303; Indels 0; Gaps 0;		
QY	787 gacaaattccagatgctgggtttcccgaaagaaatggtgcccgcggaatcgagatg 846		
DB	1 GACAAATTCACAGAGTGTCTGTTCGCGCAAGAATAATGCTCCCAAAAAGTCCGACGTT 60		
QY	847 ggtgaattcccggttcgcgatcgtaatgcagtggtgtcaaaattttaacatgaagacttt 906		
DB	61 GCGAGTTCACAGTCCCTGATCCATCCGCTTGTCAAGAACTTCAACATGGCTGATTTC 120		
QY	907 agtgggaagtgtatatacaaatggtttaaatcctacatttgaatcattgattgctcaa 966		
DB	121 AACGCAAGTGGTATATTCAAGTGGCTCAATCCCACTTTCGACACATTTCGATTGCCAA 180		
QY	967 ctctcgaatttcataaggaaaatgataaacttgttggaacttaacttaacatggcgcatataaa 1026		
DB	181 CTTCACAGATTCGCTGTCGAGGAGACAAACTATACGAACTTGATCGAGATTCGATTCG 240		
QY	1027 actttggatggtgtttcttactcgtatgctgtgcaaacatttgttcaagatccagat 1086		
DB	241 ACCCCGACTCTGGCTTCTTCCACAGAACAGCCATACAGCGGTTGTGCGAGGACCCAGCA 300		
QY	1087 ctctcgtgagcactttaataatcatgacaatgagtttcttcaactaccaagatgactggtac 1146		
DB	301 CAACCGGATCTCTATAACCATGACACAGAGTTCCTGCACATATCAAGATGACTGGTAC 360		
QY	1147 atattatcttcccaaatcgaaaaaacaccccgatgattacattatctgtatatactaccaggt 1206		
DB	361 ATTATCTCATCAAAAGTAGAACAAGAAAGATGACTACATATTTTCGTATATACTACCGCGC 420		
QY	1207 cgaacacgcatggatggatcaggtgggtccgtgaltctacacccgaaacccgacactc 1266		
DB	421 AGAAATGATGCGTGGGATGGCTATGGTGGTCTGTTCTGTACACCAACGACGAAGTTGTG 480		
QY	850 gaattcccggttcggatgtaatgacgtggttccaataattttaacatgaagactttagt 909		
DB	624 GAATTTCTCCGACACCCCTTGTCTGTCTGTACGAACTTCAACATTCGAGACTTTAAC 683		
QY	910 ggaagtggatataacaagtgtttaatcctacattgagtcattgattggtcaactt 969		
DB	684 GGAAGTGGTACATTACAGTGCTTGAATCCAACTTTGATGCCTCGACTGCCAGCTG 743		
QY	970 catgagtttcataaggaaatg---ataaactgttgggaacttaacatgagcgcaataaa 1026		
DB	744 CATGAGTTCACACAGAAGGTGACAAAGCTTGTGGAAACATCTCTTGGAGAATAAAG 803		
QY	1027 acttggatggtgtttcttactcgtatgctgtgcaaacatttgttcaagatccagat 1086		
DB	804 ACCCTAGACGTGGATTTCTTACTAGTTCAGCCGTCAAAAATTCGTGCAAGATCCTAAC 863		
QY	1087 ctctcgtgagcactttaataatcatgacaatgagtttcttcaactaccaagatgactggtac 1146		
DB	864 CAACCTGGTCTCTCAATCATGACACAGCTACTCTTCACTATCAAGATGACTGGTAT 923		
QY	1147 atattatcttcccaaatcgaaaaaacccgagattacatttctatatactaccagagt 1206		
DB	924 ATCTGTCTCAAAAGATAGAGATAAACCTTGAAGACTATATATTTGTATACTACCGTGG 983		
QY	1207 cgaacacgcatggatgatacgtggtgtccgtgctatctacaccccgacccgacactc 1266		
DB	984 GMAAACGATGCTGGATGGATGTGTGTGTCAGTTGTATACACGAGAAGTTCTGTATTA 1043		
QY	1267 cccgaatcgtatcccaaacctcaaaaagcagccaaatccgtgggtcgagactttaac 1326		
DB	1044 CCCAATACATTTATACCAAGACTCGAAAAGACGACAAAAGCATAGCCAGAGACTTCAG 1103		
QY	1327 aatttcatacaaacccgacaatagttgtggcctgagcctccatggtggaagcgttgag 1386		
DB	1104 ACATTCATTAGAACGGATAACACATGTGGTCTGAACCTCGCTGGGAGAAATTGAG 1163		
QY	1387 aaaaacgaggaagggcgagaagtgttgataaaagaactgtagagatagaagaag 1446		
DB	1164 AAGACAGTGGGAAGAGGTGAAAGGATATCTGTAAGAGGTTGNAGAGATAGAGAAGAG 1223		
QY	1447 gttgaaaagaggttgagaaggttagagatactgagatgactttgttcagaggttgctt 1506		
DB	1224 GTAGAGAAGGAAGTGGAAGGTCCGTAGGACTCGATGATGACCTTGTCCAGAGATTGGCT 1283		
QY	1507 gaaggggttaaggaattgcacaagatgaagagaatttttggaggaattgagtaagaa 1566		
DB	1284 GAAGGATTTAATGAATGAACAGACAGAGAGAGAAATTCGTGAGAGAGTTAAGTAAGAA 1343		
QY	1567 gagaaggaaattcgaatgaactcaaatgaaagcgcgactgaagtgaagaaacttttggg 1626		
DB	1344 GAGATGGAGTTTTCGATGAGATCAAAATGGAACGAAGTGAAGTTGAAAAATTGTTGGG 1403		
QY	1627 cgcgcgttaccgattgagaaacttagataaa 1657		
DB	1404 AAAGCTTTGCCAATCAGGAAGTCAAGTAGA 1434		
RESULT	5		
AF288196	1173 bp mRNA	PLN	18-SEP-2000
LOCUS	Oryza sativa subsp. indica violaxanthin de-epoxidase mRNA, partial cds.		
DEFINITION	AF288196.2 GI:10181227		
ACCESSION	AF288196		
VERSION	1		
KEYWORDS	Oryza sativa subsp. indica.		
SOURCE	Oryza sativa		


```

/translation="MSSCLLPQKCPDPSFSIHFRSFCAPKHNKGSVFVFPQCAVST
SPALLTMDVAKLRPSDPTDSLSIDROWTYTRPDGPSTEAKYLEALASSETLLTS
DEAVVVAEEAVALARAATKAVKADATLTKNSNTNLLTSADTKRSKWDOPTKERA
GILKLVASDNGELVADKLTASANKESIGDLESEKOEVEELLEOPSVSLAVRSTROT
ERKARAGLEKTASGIPSVKSGSPKKRLVAQEVHDNDPLRYLRMTSSSKLLTVR
EHEUSAGIOLDLKLERIOTELTSGROPTFAQASAAAGVDQKSLRORIHHTGLCKD
KMTKSNIRLVI SIANKYOGAMNLOVQEGCGRLVRAEKFATDGFKEFTSIYAHWI
KQAVKSLSDQSMIRLPHMYEATYRVKAEARKOLYSETGHPKNEEAEAEATGLSMKR
LMAVLLSPKPSLDOKIGMNOQLKPSVIAIDPEAVTSIDILIKBFMRQDLDKVLDLSL
GTEKQVIRWRGEMGRMKTLEIGEMVMGVSRRVRQIESAFRKLKKNKRNNHQQ
YLVAQS"
join(5386..5601,5679..5845,6111..6207,6312..6706,
6798..7311)
/note="similar to viololaxanthin de-epoxidase
emb|CAB59211.1; similar to ESTs gb|A1994713.1,
dbj|AV440942.1, and dbj|AV521985.1"
/codon_start=1
/evidence=not_experimental
/product="F22013.3"
/protein_id="AAF99753.1"
/db_xref="GI:9802551"
/translation="MAVATHCTSPCHDRIRFFSDGIGRLGITRKRINGTFLKIL
PPTOSADLRRTGSRSPRLSAFRSGFSGIFDIVPLPSKNELKELTAPLLKLVGLA
CARLIYPSADVDALATACLLKGRICELAKCIANPACAAVACLTQCNRRDETECO
IKGDLFENSVDFNECAVSRKCKVPSDLGEPAPDPSPVLQNFNTSDPENGWYI
TSLNTPDFAFQCLHEFHTEGDKLVGNISWRIKTLDSGFEFTRSAVOKFVDDPNQPG
VLYNHONEYLHYODDMTLLSSKENPEDIIFYVYKGRNDANDGEGGAVFYVRSVLP
NSIPELEKAKSIGHDFSTFTRDNTCGPEPALVERIEKTEVEEGEIRILVKEVELEE
EVEKEVKGRTENTLFORLAEGFNLKODENFVRELSEEMEFLDEIKMEASEVEK
LFGKALPIRKVR"
complement(join(7702..7832,7895..8606))
/note="similar to s-syntaxin-like protein
gb|AA00648.1|AC009540.25; similar to ESTs
dbj|AV556414.1, and dbj|AV555627.1"
/codon_start=1
/evidence=not_experimental
/product="F22013.4"
/protein_id="AAF99783.1"
/db_xref="GI:9802581"
/translation="MSVVDLKAAMKDMEAGPDFLEMASTKADKMDENLSFLEAE
YVKAEMGLISETLARIQYHEESGVHKAESVKSRLNKLISNEIVSGLRKAKTSKSLK
EMDKANKEIKRLSGTVYKRSATVGLRKLKLEVMFQGLQKRMSEKYEKTERRY
FTVTGERHDEMTEKIIDNAGGEELFTRAIQHGKGVLETVVEVLEIIDLIDRAKELEK
SLBLHGVFLDMVAVMSQANELKTAHSQRNKKWCIGIIVLLIILIVFIPIITS
FSSS"
join(10982..11218,11470..11646,11732..11896,11987..12235)
/note="similar to thioredoxin-like 1
gb|AAD35005.1|AF144387.1; similar to ESTs gb|T46281.1,
dbj|AV442467.1, and dbj|AV549179.1"
/codon_start=1
/evidence=not_experimental
/product="F22013.5"
/protein_id="AAF99754.1"
/db_xref="GI:9802552"
/translation="MTEVISTKSLFLCAGCNHHRVDDFSFSPVSGFGFLKKSFSCLK
LKSQLRSPVFGIIVFGSDSFRRSSAITAQTTLRIGTACKWKEGKLKDMREI
SSAQELVDSLTNAGDKLVVDFSPGCGGKALHPKTCQFAEMNDPVQFLOVAYEEHK
SMCYSLGVHVLFFRFYRGQGRVCSFCTNATIKKPRDALAKHDPDRCSLGPTKLE
EKELVALAANKELNFTVTPKVPVPEKEATPDSPNPLPVPSPMSSNDEKTLVSAGR"
complement(join(13219..13280,13397..13477,13866..13932,
14113..14274))
/note="unknown protein; similar to EST dbj|AV530752.1"
/codon_start=1
/evidence=not_experimental
/product="F22013.6"
/protein_id="AAF99782.1"
/db_xref="GI:9802580"
/translation="MAKYNELAKKKREAKDRKRAIHGDPLTNKLTTRTPVPSVSKR
QRKLLRWRRQKQDMVEKGLVIMEDVEMASAQAASESKSPKFSVKKSLKLNKLN
KGKKLNKQKASGESKADCMLE"
16782..19871
/note="similar to receptor protein kinase-like protein
emb|CAB66905.1; similar to ESTs dbj|AV522839.1,

```

```

dbj|AV528375.1, and dbj|AV541093.1"
/codon_start=1
/evidence=not_experimental
/product="F22013.7"
/protein_id="AAF99755.1"
/db_xref="GI:9802553"
/translation="MATPRLEFFYYIGFALFPFVSSETFQNSEQEILLAFKSDLFDP
SNNIQDWKRRENATTFSELVRLCHWTVGHCDANGYVAKLLLSNMNLSGNVSDQISFSP
LQADLNSNAFESLSKLSNLISLKVIVDSVNSFFCTFPYGLGMATGLTHVANSNN
FSGFLPEDLNATTVLEIDFRGGFEGSVSSFNKLNAFLGLSGNNGFEGKPKYVIG
ELSLLETIILGYNGFMEIPEEFKGLRLOLDLAVNLGTQIPSSJGUKQLTIVIL
YONRLTKGLPRELGCTSLVFLDLSNQITGEIPMEYVELKNLQLLNLRNQUTGIIP
SKIAELPNLEVLWQNSLMGSLPVLHKGKNSPLKLDVSNKLSGDIIPSGICYSRNL

```

Query Match	Score	DB	Length
Best Local Similarity	66.38	Pred. No. 6.6e-76	
Matches	668	Mismatches	246
Indels	94	Gaps	2

```

QY 744 atgtcagataaaatgtgtgacttgttcgaaaaaacagtggtgagaccatccaacgagtg 803
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6305 ATATCAGATTAAATCTGGGGATCTGTTTGAGACACAGTGTGTTGATGAGTTCAACGAGTG 6364
QY 804 tgcggtttcccgaaagaaatgtgtcccccgaatcggatgtgggtgaatcccggttcc 863
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6365 TCGTGTGTCGAGAAAAAGTGTGTTCTTAGAAAAATCTGATCGGAGAAATTCCTGCCCC 6424
QY 864 ggcgtgaatgcaglggttcacaaattttaacatgaaagactttagtggaagtgtatat 923
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6425 AGACCCCTCTGTTCTGTACAGAACTTCAACATCTCGGACITTAACGGGGAAGTGTATAC 6484
QY 924 aacaagtggtttaaatcctacatttgatgcatttgatgtcaacttcagatttcacat 983
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6485 TACAAGTGGCTTGAATCCAAACCTTGTATGCTTCGACTGCCAGCTGCATGATGTTCCACAC 6544
QY 984 gaaaaatg---ataaactgttgggaacttaacatggcgacataaaaaactttggatgggtg 1040
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6545 AGAAGGTGACACAACTGTTGGAAACATCTCTTGGAGATAAAGACCTTAGACAGTGG 6604
QY 1041 ttcttactcagatctgtgtgcacacattttcaagatccagatctctcctggagacat 1100
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6605 AITCTTTACTAGTACAGCGTACAAAAATCTGTCGAAGATCTCAACCAACCTGGTGTCT 6664
QY 1101 ttataatcagcaatgactttcttcacaccacaaatgact----- 1141
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6665 CTACAATATGACACAGAGTACCTTCACATCAAGATGACTGGTAAATACATCTTGATC 6724
QY 1142 ----- 1141
DB 6725 CACCCACGAGGAGACTTTGATAGTGTCTATTCTTTTGGCTTAAACAGCTTTTCAT 6784
QY 1142 -----ggtacatatattcttcccaaatcgaaacaaaccccgatgattacat 1189
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6785 ATGATTTTTCAGGTATATCTCTCAACAAAGATAGAGAAATAACTGAAGACTATATAT 6844
QY 1190 tctatactaccaggtcgaaacgacgcatggatgatacagtggtggtgcgtgatctaca 1249
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6845 TTGTATCTACCTGGCGGNAACGATGCTTGGATGATATGTTGTCGATGTTGATACA 6904
QY 1250 cccgagccgcgacactcccggaatcgatccatcccaacacatacaaaagcagcaaatccg 1309
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6905 CGAAGACTCTGTATTACCCCAATAGCATTATACCAAACTCGAAAAAGCAGCAAAAAGCA 6964
QY 1310 tgggtcgagactttaacaatttccataaacacgacacatggtgtggtggtggtggtcct 1369
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6965 TAGGAGAGACTTTCAGCACATCTATTAGAACGGATACACATGTTGCTCTGACCTGCGC 7024
QY 1370 tgggtgaaagcgttgagaaaaacagcggaagggcgagaggttggatgataaaagaagctg 1429
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7025 TCGTGGAGAGAAATTGAGAAGACAGTGTGAAGAGTAAAGGATAAATCGTAAAGAGGTTG 7084
QY 1430 tagagatagaagaagcgttgaaaaagaggtggagaaggttagagatactagagatgactt 1489
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7085 AAGAGATAGAAGAAGGTTAGAGAAGAGTGGAGAGGTCGTTAGACTGTAGATGACCT 7144

```


Db 190 VCCNNWKRMAHHHWCATNNNNWWYAYIMHHKKGAATNNKTBARDDBAHVKY 131
 QY 1000 gttgggaacttaacatgctgcgcataaaactttg 1032
 Db 130 WWRDYWCACWNNNAKAYRTAKHMYVTDR 98

 RESULT 11
 AY008297
 LOCUS 4788 bp mRNA ROD 10-APR-2001
 DEFINITION Mus musculus RPRG-interacting protein mRNA, complete cds.
 ACCESSION AY008297
 VERSION GI:11496166
 KEYWORDS
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostei
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 1 (bases 1 to 4788)
 Hong,D.H., Yue,G., Adamian,M. and Li,T.
 Retinitis Pigmentosa GTPase Regulator (RPRG)-interacting Protein
 Stably Associated with the Photoreceptor Ciliary Axoneme and
 Anchors RPRG to the Connecting Cilium
 J. Biol. Chem. 276 (15), 12091-12099 (2001)
 J1104772
 2 (bases 1 to 4788)
 Hong,D., Yue,G., Scimeca,M.S. and Li,T.
 Direct Submission
 Submitted (26-SEP-2000) Berman-Gund Laboratory, Harvard Medical
 School, Massachusetts Eye & Ear Infirmary, 243 Charles Street
 Boston, MA 02114, USA
 FEATURES
 source Location/Qualifiers
 1..4788
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 79..4074
 /note="RPRGIP"
 /codon_start=1
 /product="RPRG-interacting protein"
 /protein_id="AAG22857.1"
 /db_xref="GI:11496167"
 /translation="MOHLLEYMPEDLPVRDTSPLLKGTSGKNVRAQHPLGLG
 LNCRLHLEPTVLPKPSQKQDKNRRRTNVQSTTTPDLRTLVAYLQPE
 WYASPSAPPRAPGRKAHVRLCPSTAVGSAQPRKHGRPLPHKVP
 APFAFKVADKNTRILHTEPQSLTHTTSTVHEELPRSPKPTSAVEP
 ECTQAAERASIRINVELRLKRLQERNLTLAETAGLTVEQAEVLDLQ
 DTAHAFILSQYNELKAELESSEKSKAVSLTQLGDSVILGILTKLEQVREVDL
 LSDSYDRLENNLSDSHQPSHWSTLTGQLPQVCPQLDDQMGTLA
 ROATNKAQDGLKFOQTDILYHQEEFESLQATVASSPEELCEIAAQPY
 QRESPEAKQNDLSVLQYVSHAEITTELEKTRDMLLQRLKINCMYQE
 TKADRENDRHEKLERLHLLDFKNSRIKLEGILRSHGLPTSEQLKDVAY
 CLEPLAHRGDEVDMSLLHPSENLELHVHOAFLTPAALTOAGDTQPTTFC
 FTHCTPQGPQLDFTSQYVQADYLLHVLQTSVRLDLHQAMASEVYR
 ISDLKVLGTVEHGLATLAGAGEDGLVLEWRLCLKPLKSLGCAKNRKK
 VSVLGAQRQVNEGRSETWAPQNELRVETRCGLRSRLRHQPSQYVWIRAF
 DTIILPASNSYFKDQALFVITLDQYLRREALSVYFDDDEPEPGSYL
 LPLAQNKIGDRNLDTSGKSGSIKVLQDMKSHYLAPEGFQMSAEKPE
 EGGEVEKEEVEEEDNDVLEASTFEWVPFFSQDOIASTEIP IEAGQVPEK
 EKKEREHOVASIRKHSKPGQVDKRNMEYLSNLLNGNTQOMHYTEMK
 ABEKREHOVASIRKHSKPGQVDKRNMEYLSNLLNGNTQOMHYTEMK
 EDGGLKQVADKREEPSKSLALRQEHPSHPNATSLADQSCFQASEVSETOT
 IYTPQAQYVPKADSPKMCIEIVSLAFCEADVMSDETIOOVVEYVKFCDLP
 MSLRPRAGEIEHFHFSSVLDLPVEHQSRQFLFAMLHAQSDSEGRKFTV
 EKKFCODIGYAYLELWQIFQSGKDILLEQEIVSPRNQAIQIGRLKLSQAG
 IYKMETDLS"
 BASE COUNT 1442 a 1156 c 1253 g 937 t
 ORIGIN


```

Matches 133; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

Qy 1356 gcttgagcctccattgtggaagccttgagaaacacggaagagcgagaggttgtt 1415
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2781 GTCAGAGCTGAGAACGCCGAGGCTGAAGAGAGAGGAGGAGGAGGAGGAGGT 2840

Qy 1416 gataaagaagctgtagatagaagaaggttgtaaaagaggttgagaggttagaga 1475
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2841 AAAGGAGGAGAGGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2900

Qy 1476 tactgagatgacttcttgcagaggttgcagaggttgaagggttcaacaagatga 1535
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2901 AGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 2960

Qy 1536 agagaatttctgaggaggttgagaaagagagagagagagagagagagagagag 1595
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2961 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 3020

Qy 1596 ggaagcgcactgaagttgaaaa 1616
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3021 GGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 3041

RESULT 12
AC079533/c
LOCUS
DEFINITION
Mus musculus clone RP23-348M1, WORKING DRAFT SEQUENCE, 7 ordered
pieces.
ACCESSION
AC079533.1 GI:9964898
VERSION
HTG; HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 201978)
DOE Joint Genome Institute.
Direct Submission
Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Project Name: 1870793
Center Project Name: 1870793
Center clone name: RPCI-23_348M1
-----
Summary Statistics
Consensus quality: 197680 bases at least Q40
Consensus quality: 200626 bases at least Q30
Consensus quality: 201212 bases at least Q20
Estimated insert size: 205000; agarose-fp estimation
Estimated insert size: 201728; sum-of-contigs estimation
Quality coverage: 10.36 in Q20 bases; agarose-fp estimation
Quality coverage: 10.53 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 14235: contig of 14235 bp in length
* 14236 14335: gap of unknown length
* 14336 40527: contig of 26192 bp in length

```

FEATURES

source

1..201978

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="RP23-348M1"

/clone.lib="RPCI mouse BAC library 23"

55239 a 43169 c 44386 g 58583 t 601 Others

ORIGIN

```

Query Match      2.8%; Score 56.2; DB 76; Length 201978;
Best Local Similarity 51.0%; Pred. No. 0.099;
Matches 133; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

Qy 1356 gcttgagcctccattgtggaagccttgagaaacacggaagagcgagaggttgtt 1415
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60839 GTCAGAGCTGAGAACGCCGAGGCTGAAGAGAGAGGAGGAGGAGGAGGAGGT 60780

Qy 1416 gataaagaagctgtagatagaagaaggttgtaaaagaggttgagagaggttagaga 1475
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60779 AAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 60720

Qy 1476 tactgagatgacttcttgcagaggttgcagaggttgaagggttcaacaagatga 1535
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60719 AGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 60660

Qy 1536 agagaatttctgaggaggttgagaaagagagagagagagagagagagagagag 1595
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60659 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 60600

Qy 1596 ggaagcgcactgaagttgaaaa 1616
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60599 GGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 60579

```

RESULT 13

AC074378/c

LOCUS

DEFINITION

Homo sapiens chromosome 2 clone RP11-4117, WORKING DRAFT SEQUENCE,

10 unordered pieces.

ACCESSION

AC074378

VERSION

AC074378.3 GI:9929770

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

WATERSTON, R.H.

The sequence of Homo sapiens clone

Unpublished

2 (bases 1 to 156610)

WATERSTON, R.H.

Direct Submission

Submitted (30-JUL-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

On Aug 26, 2000 this sequence version replaced gi:9755476.

COMMENT

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc/index.shtml


```

CDS
weak"
4345. >5115
/genes="611.8"
/note="contains zinc finger domain, contains similarity to
deor family (bacterial regulatory proteins)"
/codon_start=1
/product="611.8"
/protein_id="AAF77208.1"
/db_xref="GI:8573035"
/translation="MSQQWTFDQRRQIAEMLRNKGVICRDLFTGRCSPPTCPYM
HVGANGRPVMSVCTFTTQKLRDGGFFHGTAQLEELHAGAPVRYRQYMKIA
VPPAEVNLGDIATHLSIAAFYANPMPALHVVHGPTVSHENAVNSFQPLVMPQNSQAI
APTMEASQGTASTPSPHFAFYANAPITSMPTHTSAATVLOPHTLVQPTTFYQANA
LLAQQPQHFAPVQPLNTQQHQLTQALGSRVYFHIQPH"
11016 12407
/genes="611.9"
/note="predicted using Glimmer, Testcode and CodonUsage"
11016. >12407
/genes="611.9"
/note="contains ATP/GTP binding domains, has similarity to
Gnpases"
/codon_start=1
/product="611.9"
/protein_id="AAF77209.1"
/db_xref="GI:8573036"
/translation="NRATLRQCNVDFAPMRKVKRTLSAYKGSMTSAAKSMSEAFV
DIDEGAAYLGHMQSAANLADVKDADFLEIRDARLPTTENPNRKLTAQKPRL
IIFNKALSNEDSNRAIOEYERNCAFALFTSARCARDVVEAORFTHLPLPYK
TVHAGLVGMNVGKSLINSLRLAHEQFHREDFRSPSPETVSTPTGKMLV
PUKDPVVDYDTPGLTLPFGCTKESGLKLAACGIPTNDSLPQGMVARYIYDILVA
SGSSHAECLHLPRVPSFDVCVAMICERSGTSGTGMGLDPRVRAHREVFHDFIMG
NLGKTLTDVLRLLRSOSSIPLESSTAAATDGAQAAGDDGAYATHVETSDV
VERYPDHMRDVLQSLKPLHHDGVPRRHQAFSVKNLDSFSDCTVISRKKGISRTAF
DQLAHRHTLPRGR"
13052. >14005
/genes="611.10"
/note="predicted using Glimmer, Testcode and CodonUsage,
Glimmer predicts second ATG as start codon"
13052. >14005
/genes="611.10"
/codon_start=1
/product="611.10"
/protein_id="AAF77210.1"
/db_xref="GI:8573037"
/translation="MANASAAASDELRTQOYLVEIDECVFERORELAPRRLLSSOL
ETOKTLVNDLFWLHDVCRYTSTIEDATLPALETLPASSPARQQLQSRNMYRDLAT
RLCRGGRDSDVPLLATPRAFSGSRAADVPLISHDSAAAGAHVQWRLLEHARLL
DALARATARIETHVAETAENGELAUARMNELTRVYVYSGSMWLKPPPEASAAQPP
EGVSATARFAEARALAYEKTQVQALNNELALLHNNALSRARTREYDMLKRHMAEDQ
RKHEDQIAECDAVILGRMSVELEQLIQENALKHLRMTAELD"
14801. >15382
/genes="611.11"
/note="predicted using Glimmer, Testcode and CodonUsage"
14801. >15382
/genes="611.11"
/note="contains S-adenosylmethionine-dependent
methyltransferase domain, has Blastp similarity to
hypothetical protein W02B12.10 from Caenorhabditis
elegans"
/codon_start=1
/product="611.11"
/protein_id="AAF77211.1"
/db_xref="GI:8573038"
/translation="MVNPVVFVDVGCAGFMFLSAVFPATCMGLRTPKVVFA
QEKALAROENASLTOHYRNWFEQNMVKFSNCTKQGLSKLFFCPDPHWRKN
IRRLIISGLVQETAYLWRGLGLVYTVSDVAELFEMVMVQCLDSDPLFRRLAEALCTP
EHKQVLEIVTGASEDAQRTARKGLQKHFAVHMRT"
17686. >18489
/genes="611.12"
/note="predicted using Glimmer, Testcode and CodonUsage,
Testcode weak in 5' end"
17686. >18489
/genes="611.12"

gene
/notes="contains ALDEHYDE_DEHYDR_CYS Aldehyde
dehydrogenases cysteine active site (from Prosite)"
/codon_start=1
/product="611.12"
/protein_id="AAF77212.1"
/db_xref="GI:8573039"
/translation="MHLFALPLLSVLAVLAVPGRGASKCFVTADTPTVPFGLELCY
MHSNACCLFPGNDKIDILTAYSAIVPKQGVAGSQRIYTSILYALRQVLCICLPCKEPL
YRFSVAGDIDVGLGVNSTAGETQWTRICRSLYKQKTRKGLWSDGSRVACGG
VIVSSCMSTPVENITAASFQSPSPCVASNELIIPSVTRGSPNPALEMLSMWTQSTP
DFQIIVINDSPDNYDKTCFCGRDATATVASAALSXWVSAVALALLLWL"
20647. >22497
/genes="L165.1"
/note="predicted using Glimmer, Testcode, CodonUsage"
20647. >22497
/genes="L165.1"
/note="contains COG0438 Glycosyltransferases 1, contains
prosite profile Prokar_Lipoproline and Prosite pattern
Prokaryotic membrane protein, contains Pfam Glycos Transf
1 Glycosyl transferase"
/codon_start=1
/product="L165.1"
/protein_id="AAF77213.1"
/db_xref="GI:8573040"
/translation="MLMCECTFASLCNCPDPRLGALYCLRLMGYACVRDPARPS
VEAVRVRADDOQPARKACFLSFCGWSARGLDSVATPLQVSCVAAQTHRLDLSLD
TMATPPSPPTSLAHLSSIPRYFVSHQHRRTAGAVYMTTCGRVIRMLTPTLVLL
AALALCHCKRAYARTNTQFLHAAAGAGGGERVLMVALDGLQADAAARGVKQYVL
FTNEKPADLSAESDHLSELVEKQFSTRLLRPVRFVILRPALTRWLSGDYPRLT
LLIQTKGGAALFEVAVANATPIVETVGVFPFAPLLRLLAGCMVSYTHYPIVSS
AMTORVSGEVSHNTSTVAVNMLCAKVIYGVFSLTRCMGFPFNVLTNSWTQ
NHQVIFWRACIRLPPCDVAGFAAGQPPALNNRIVSGQFRPEKNHMLQLVAFH
AMPRPLRDAKLVMIGARNADKRAEQHLVRAKELGIEQVQLLVNATVAEVQAEEL
GKCVIGLHTMRDEHFGVILLEYLAAGCIPLGRSGGVDELILNSDGLFLAVTAEBYA
AAMVEICEMLRDPRDQRYVQFKRGSEHVKSFDSDSPTRFVELVSEVYAC"
23148. >23402
/genes="L165.2"
/note="predicted using Glimmer, Testcode, CodonUsage"
23148. >23402
/genes="L165.2"
/note="contains Gribkov GINA_1 Glutamine synthetase
signature 1"
/codon_start=1
/product="L165.2"
/protein_id="AAF77214.1"
/db_xref="GI:8573041"
/translation="MSYRSEAKKEEFKRYLESTQVVDALTRVLVNLVEEKEPDPV
DYIKRVLGGASSADYEALQENARLRAEVELLKKLGGQAQ"
24858. >25517
/genes="L165.3"
/note="predicted using Glimmer, Testcode, CodonUsage"
24858. >25517
/genes="L165.3"
/note="contains WD40 repeats"
/codon_start=1
/product="L165.3"
/protein_id="AAF77215.1"
/db_xref="GI:8573042"
/translation="MKRVFDSAKPNDPFCSPVTSCDWNSEDINTVACCSIDSTVTLWDV
ETGAOKTKLVADKVDYDIAFAASHTFASGADGSRVFFDLRMDHCTILYETQGLSP
LLRLAWNQDFDVFATFGIDSDPAVVDIMRYPTVPASQLSQLHQLPINLWSPQNAQ
NICTAGEDGLVCVWEAREKGRSILWCDCVEFINNVARRAQNADWMAITTSKGAQLL
PL"
26533. >27297
/genes="L165.4"
/note="predicted using Glimmer, Testcode, CodonUsage"
26533. >27297
/genes="L165.4"
/codon_start=1
/product="L165.4"
/protein_id="AAF77216.1"
/db_xref="GI:8573043"
/translation="NSLLNSTLQTLVWRLRDMSGNVTOQKLNHRVFDAYEAKSLVFQA

```

```

ISPAQQLVMKQYSGRIIPMPHPLGQPLMVDSWSELVHLKPDNEVOLLPRRARNNNAYA
VMSATCCSAGSPFEMDHRLEPDKLVKFSQADQDARTAFNLKHTDKVPQTIFLDGLM
EAPRASALVSHFNILTPAHVNNLAGTEOFLREWCRAPADGDRHRQLKCFSSLLEKOT
HLFLGTNAAPGCRELLNYAKGNIIFYAKKGMAFYVP"
27619. 27753
/gene="L165.5"

Query Match      2.8%; Score 55; DB 4; Length 39987;
Best Local Similarity 51.4%; Pred. No. 0.16;
Matches 127; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 1371 ggtggaaggcttgagaaacagcggaagagcggaaggttggttgataaagaagcctgt 1430
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2678 GGAGGAACCAAGCGCGGATGTGGGAAGAAGAGGAGGAGGAGGAGGAGGAAGA 2619
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1431 agagatagaagaagaggttgaaaaagaggtggagaaggttagagatactgagatgacttt 1490
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2618 AGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGA 2559
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1491 gtttcagaggttgcttgaaaggggttaagagaggttgcaacaagatgaagagaatttgtgag 1550
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2558 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGA 2499
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1551 ggaagttagtaaaagagaagaagaattctgaatgaacttcaaatggaagcgactgaagt 1610
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2498 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAAGA 2439
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1611 tgaagaag 1617
      || ||
Db 2438 GGAGGAG 2432

RESULT 15
AC005804/c
LOCUS
DEFINITION
Leishmania major chromosome 35 clone 6.1.1 strain Friedlin,
complete sequence.
AC005804
VERSION
AC005804.2 GI:6006841
KEYWORDS
HTG.
SOURCE
Leishmania major.
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE
1 (bases 1 to 42160)
Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,
Cawthra,J., Marsolini,F., McDonagh,P., Stuart,K.D., Cunningham,M.
and Beverley,S.
Direct Submission
Submitted (09-OCT-1998) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
REFERENCE
2 (bases 1 to 42160)
Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,
Cawthra,J., Marsolini,F., McDonagh,P., Stuart,K.D., Cunningham,M.
and Beverley,S.
Direct Submission
Submitted (02-OCT-1999) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
REFERENCE
3 (bases 1 to 42160)
Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,
Cawthra,J., Marsolini,F., McDonagh,P., Stuart,K.D., Cunningham,M.
and Beverley,S.
Direct Submission
Submitted (31-MAY-2000) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
COMMENT
On Oct 2, 1999 this sequence version replaced gi:3723957.
FEATURES
Location/Qualifiers
1..42160
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="35"
/clone="6.1.1"

```

```

/note="Also known as 611"
535. .>2694
/gene="511.9"
535. .>2694
/gene="511.9"
/note="contains a proline-rich region (PRO_RICH),
contains reca family and SURE1 domains (Pfam)"
/codon_start=1
/product="511.9"
/protein_id="AAF73077.1"
/db_xref="GI:8122319"
/translation="MMAVAQLEEQNLNDLALNRASAKAIDSPADTSSAISAARANEIIR
DKPMDVAQENITLTTEMAKARCYHPPIIPAVLTGKESSEWAIARCMYLVYADD
YLASHPNHYGTFEIARTEMOLQIIVRISEDFVKSAFINAVORSKVPSPGADGSTG
QAGSLSVAPRMDSALTAPTANRLIRNEAALOGVDIIVORLGENENRTDILKDVRLK
LEKLVRAVQAQPDGAYRDEETGRGPTLSRSLLPVLYKHQHPHLRVSKASARELY
TDACALRYVLTPLDDIADAVDIPSELATVVFELVYRAIKVIOLDTSFFADPKLF
VDSRGQIGLFTPRYRDIYIFGLDFMEEFWRKMLSKSIPIGENHDFIDHVDSDVN
FYIRIELLDGYVNAKGALEKESKEYARSMRTEWFPSCVDCVHESTNVLYFHKTL
LTFYGLRIVLYGSGVIGADYEQCEVEDYITRGVAGVDDQLVLAFAAELQOEA
LAKRNHTKSNLNLTSVRLSDVDFMLNNLCFLDNRYRREACTFRLAAETPPPKS
SCGASARBDTAAPPPDPPLSOLFDMINSERERYVEAFGASWHKCFPIKIGOSDLS
IEFTWELKKSORMAYKOWHRVNDALLKLYCKAEAMDAKSKTQFLDVSVAARV
DGFERMEVLLNGRTWSQFQKMSMSTPTTWEQDQIRQAF"
3675. .>4784
/gene="511.10"
3675. .>4784
/gene="511.10"
/codon_start=1
/product="511.10"
/protein_id="AAF73078.1"
/db_xref="GI:8122320"
/translation="MERNPAHPLDATALKIPVROKEVTFDSLVCVNADRFEVGLSGKSE
ERRGFRNTKLYAANAARTGRGTGASDLGFTESLPTLNEVIGEEERKISHGLRLRARP
RAAGSTVDTPVAVKAKSLTRPVHATPTISPTSLRWGSSVQNVQGHASAAAT
HLSDGVAERLRAKVEAARLAKIRALRAVVKVQMOSFIDRGSAFASITVGVKMSDT
LAVARCSVINTLVGTQVQNRARFRRLCAARGPVSGQLFPNINAGQSDRFTLSLY
SSAPGARESSVNESYSPNSVDYLEEEVSRIKOLYDFDYFCEAVSISAVASGSD
MSIFAVSAASSISRPSPSPVSRGDDHSHSFHCSTTM"
6016. .>6735
/gene="511.11"
6016. .>6735
/gene="511.11"
/note="contains SAM-dependent methyltransferase domain"
/codon_start=1
/product="511.11"
/protein_id="AAF73079.1"
/db_xref="GI:8122321"
/translation="MNHVEWFMQYAMVEAALKACLRAPTVLSKDGATRIILHMGCGNS
DFCDHVBGLSLDHPVPSSSSRASEVINVDICESIVAHLAHPSPRLYAVGCCDLHV
SSSPMPFPSSNAWYSPDRLRLKVLQSSVDVDFDKGTADALLSFAGYNPNMEAY
MGEALKVRPGCLLFLISINSEDLSPYALSADDGLKSFHLAYTDVIELGAHLRLHLR
VETLGSYSCYGYAVVASAAVE"
7085. .>8977
/gene="BT1"
7085. .>8977
/gene="BT1"
/note="contains similarity to T. brucei ESAG10, contains
transporter motifs and multiple transmembrane domains"
/codon_start=1
/product="Bioprotein Transporter 1"
/protein_id="AAF73080.1"
/db_xref="GI:8122322"
/translation="MPGRYRVPSAPESPPOGAKYVHPVSAHLVRAAPFLGYPVFSI
VIRFHPKVLAICIORLFEKGLADLRMSIOPMLTGRYGLTGCMYORISLYTLGW
VIRFHPKVLAICIORLFEKGLADLRMSIOPMLTGRYGLTGCMYORISLYTLGW
ALNFIIVMADTFALFGYTRKWCYVMSAVGSGVFLYGLLPKAKESAKPAALMFTL
ALFMSINDVFAVALYSEQIRRRPAAGPALVSWMMGTALIGIMISSVILGSLNGLAH
FGVYITAILLSGLLEFVNLFEERNRAARLEDAEMIEFLQKTKSASESAGSPSP
HKLDGHLTIDNRVEEDNDVDEQAAAADQGFVRPMDTYLCGAVSMNRDVLNRWR
MALFCLITLTCVIANSLVILGTWMDIMVCIVLAAVLCVSSFFTLPAKAVVPMY
FNAILYLSLPCVINTFYVAKPCLPDGPHFTSYFNAMNGFLGNIGIYFVTHLF
PHSYRVPMGLSALLPAASMFVDLILKRNLAIGIPDHAMVIFGDTIIEVCNMLLN
MPMMMLMCRIPRGSSEMVFPALLASIIHLGTSTSSAIGLYLIMETIWPVVTOGQCDYSN

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2001, 23:49:01 ; Search time 313.48 Seconds
(without alignments)
3967.950 Million cell updates/sec

Title: US-09-075-375A-1

Perfect score: 181
Sequence: 1 tttggttcgaatttacc.....aattttggctaaaaaaa 1981

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1894.6	95.6	1981	18	AA166241 Romaine lettuce vi
2	643.8	32.5	1589	18	AAT66242 Tobacco violaxanth
3	610.2	30.8	1555	18	AAT66243 Arabidopsis violax
4	116.6	5.9	936	22	AAF58252 Oligonucleotide D1
5	116.6	5.9	936	22	AAF58252 Oligonucleotide D1
6	116.6	5.9	936	22	AAF58252 Oligonucleotide D1
7	116.6	5.9	936	22	AAF58252 Oligonucleotide D1
8	116.6	5.9	936	22	AAF58257 Oligonucleotide D1
9	116.6	5.9	936	22	AAF58257 Oligonucleotide D1
10	116.6	5.9	936	22	AAF58259 Oligonucleotide D2
11	116.6	5.9	936	22	AAF58259 Oligonucleotide D2

c	12	116.6	5.9	936	22	AAF58262	Oligonucleotide D2
c	13	116.6	5.9	936	22	AAF58262	Oligonucleotide D2
c	14	116.6	5.9	938	22	AAF58255	Oligonucleotide D1
c	15	116.6	5.9	938	22	AAF58255	Oligonucleotide D1
c	16	60.6	3.1	244	22	AAF58238	Oligonucleotide D1
c	17	59.2	3.0	244	22	AAF58238	Oligonucleotide D1
c	18	44.8	2.3	51259	18	AXA83007	Partial mouse WRN
c	19	44.6	2.3	425	20	AXA1515	Human secreted pro
c	20	43.4	2.2	1120	21	AXA46213	Arabidopsis thalia
c	21	43.4	2.2	1123	21	AXA35071	Arabidopsis thalia
c	22	43.2	2.2	827	21	AXA47631	Arabidopsis thalia
c	23	42.8	2.2	49999	20	AAZ23891	Murine LOBO genom
c	24	42.8	2.2	49999	20	AAZ23896	Murine LOBO genom
c	25	42.6	2.2	2169	21	AAZ50687	Arabidopsis thalia
c	26	42.6	2.2	2638	21	AAZ47245	Arabidopsis thalia
c	27	42.4	2.1	3211	18	AAT89346	Human p160 cDNA 16
c	28	42.4	2.1	3501	18	AAT89345	Human p160 cDNA 16
c	29	42.2	2.1	15464	12	AAQ11415	Ryanodine receptor
c	30	41.8	2.1	5340	22	AAF26289	P. falciparum eg5
c	31	41.8	2.1	910715	20	AAQ20248	Borrelia burgdorfe
c	32	41.6	2.1	507	13	AAQ32658	LEM1 10 cDNA. Ly
c	33	41.2	2.1	1505	15	AAQ55750	Genomic clone g1f
c	34	41.2	2.1	9408	21	AAA70157	Plasmodium falcipa
c	35	41.2	2.1	9636	15	AAQ67190	P. falciparum tran
c	36	41	2.1	2601	21	AAZ46719	Arabidopsis thalia
c	37	41	2.1	1664976	19	AAV21209	Methanococcus jann
c	38	40.8	2.1	1484	18	AAV74906	Staphylococcus aur
c	39	40.8	2.1	6002	21	AAZ86968	Retinoblastoma bin
c	40	40.6	2.0	50000	21	AAA64139	Nucleotide sequenc
c	41	40.6	2.0	49999	20	AAZ23895	Murine LOBO genom
c	42	40.4	2.0	1530	21	AAZ50661	Arabidopsis thalia
c	43	39.8	2.0	456	21	AAA43753	Human secreted exp
c	44	39.6	2.0	591	20	AAZ20389	Borrelia burgdorfe
c	45	39.6	2.0	1998	21	AAA70212	Plasmodium falcipa

ALIGNMENTS

RESULT 1
AA166241
ID AAT66241 standard; cDNA; 1981 BP.
XX
AC AAT66241;
XX
DT 28-JUL-1997 (first entry)
XX
DE Romaine lettuce violaxanthin de-epoxidase cDNA.
XX
DE Violaxanthin de-epoxidase; VDE; light; photosensitivity;
KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
KW xanthophyll; lettuce; ss.
XX
OS Lactuca sativa L. cv. romaine.
XX
FH Key Location/Qualifiers
FT misc_difference 26...29
FT FT /*tag= a
FT FT /note= "bases 26-29 are illegible in Fig 1"
FT FT misc_difference 66..72
FT FT /*tag= b
FT FT /note= "bases 66-72 are illegible in Fig 1"
FT FT misc_difference 105..110
FT FT /*tag= c
FT FT /note= "bases 105-110 are illegible in Fig 1"
FT FT misc_difference 147..149
FT FT /*tag= d
FT FT /note= "bases 147-149 are illegible in Fig 1"
FT FT misc_difference 186..189
FT FT /*tag= e
FT FT /note= "bases 186-189 are illegible in Fig 1"
FT FT misc_difference 226..227
FT FT /*tag= f

```

FT /note= "bases 226-227 are illegible in Fig 1"
FT 235..1656
FT *tag= g
FT transit_peptide 235..609
FT *tag= h
FT mat_peptide 610..1653
FT *tag= i
XX W09717447-A2.
XX
XX 15-MAY-1997.
XX
XX 07-NOV-1996; 96WO-US18291.
XX
XX 06-AUG-1996; 96US-0023502.
XX
XX 07-NOV-1995; 95US-0006315.
XX
XX (CALJ ) CALGENE INC.
XX
XX Bugos RC, Rockholm DC, Yamamoto HY;
XX
XX WPI; 1997-281036/25.
XX
XX P-PSDB; AAW09874.
XX
XX DNA encoding plant violaxanthin de-epoxidase - used to modify the
XX sensitivity of a plant to light
XX
XX Claim 3; Fig 1; 41pp; English.
XX
XX A cDNA clone (AA066241) codes for the 55 kDa violaxanthin de-epoxidase
XX (VDE) (AAW09874) of romaine lettuce. VDE was purified from romaine
XX lettuce chloroplasts and 2 tryptic peptides were used to develop
XX primers (see also AA066244-45), which amplified a partial VDE
XX sequence. The amplified sequence was then used to screen a lettuce
XX cDNA library, and the 1981 bp DNA sequence was identified. VDE
XX nucleic acids (see also AA066242-43), in sense or antisense
XX orientation, can be used in genetic constructs, pref. also contg. a
XX plastid translocation sequence, to modify VDE levels in plants.
XX Increased levels result in the plant being tolerant of increased
XX light and therefore more productive and/or more resistant to
XX disease. Underexpression of VDE increases photosynthetic
XX efficiency under low light. The photosensitivity of a range of
XX crops, trees and ornamentals can be modified.
XX
XX Sequence 1981 BP; 608 A; 337 C; 433 G; 577 T; 26 other;

Query Match 95.6%; Score 1894.6; DB 18; Length 1981;
Best Local Similarity 97.4%; Pred. NO. 0;
Matches 1931; Conservative 0; Mismatches 49; Indels 2; Gaps 2;

Qy 1 tctgggttcgaatttaccaccacaaagtgttgcctaccataattgggataaggagtct 60
Db 1 tctgggttcgaatttaccaccacacacnnnttgcctaccataattgggataaggagtct 60
Qy 61 aatttccctgtacatttcccaatttctctccgcacaccataataatactgtacgc 120
Db 61 aatttccctgtacatttcccaatttctctccgcacactannnnnntactgtacgc 120
Qy 121 cacttcgaagctacaagtgttgaaagacacagcagattttacaagacgagagataa 180
Db 121 caacttgaagcgtacaagtgttgannnngacgtagattttacaagacgagagataa 180
Qy 181 taagc-ttaagctactcgatcgctcagtggtcctttggagccaaacaaactgctatggc 239
Db 181 taagcnnnaagctactcgatcgatcgctcagtggtcctttggagcc-acnnactggctatggc 239
Qy 240 tctttcttcaactgtatttctctcaagaggaagccctcaatttatcaagatc 299
Db 240 tctttcttcaactgtatttctctcaagaggaagccctcaatttatcaagatc 299
Qy 300 accatgtaagtgaagggttccacagagtgagacacacctcctaccacaataatcatgtaa 359
Db 300 accatgtaagtgaagggttccacagagtgagacacacctcctaccacaataatcatgtaa 359
Qy 360 aattcgatccaacaatggatattttaaatttccgggtgtgttacatcttataagaaga 419
Db 360 aattcgatccaacaatggatattttaaatttccgggtgtgttacatcttataagaaga 419
Qy 420 ttotttctcagattctagccattgcaaggataaattcagatcgagcattgatacaag 479
Db 420 ttotttctcagattctagccattgcaaggataaattcagatcgagcattgatacaag 479
Qy 480 ttttgaggaatacaagaattgattctcaaaagggcgatgacttcttgaataagca 539
Db 480 ttttgaggaatacaagaattgattctcaaaagggcgatgacttcttgaataagca 539
Qy 540 atggagacaattcacaattggctatcgtattgtttgacacatttcttctccag 599
Db 540 atggagacaattcacaattggctatcgtattgtttgacacatttcttctccag 599
Qy 600 agttgatcgctgtatgctcttaaaacttggctgttttactcaaaagaatgcaggattga 659
Db 600 agttgatcgctgtatgctcttaaaacttggctgttttactcaaaagaatgcaggattga 659
Qy 660 gcttgcaaaatgtatagcaaaacccatcttgcgcgcaaacgttgcctgtctacagacttg 719
Db 660 gcttgcaaaatgtatagcaaaacccatcttgcgcgcaaacgttgcctgtctacagacttg 719
Qy 720 caacaatcgtctgacgagaccgaatgtcagataaaatgtgtgacttgttcgaaaaag 779
Db 720 caacaatcgtctgacgagaccgaatgtcagataaaatgtgtgacttgttcgaaaaag 779
Qy 780 tctggtgaccacaattcaacgagtggtgcggtttccgcaaaagaataatgtgtcccggaatc 839
Db 780 tctggtgaccacaattcaacgagtggtgcggtttccgcaaaagaataatgtgtcccggaatc 839
Qy 840 ggaatggtgaaatcccggttccggaatcgttaalgagtggttcaaaattttaacatgaa 899
Db 840 ggaatggtgaaatcccggttccggaatcgttaalgagtggttcaaaattttaacatgaa 899
Qy 900 agactttagtgggaagtgtatatacaaaagtggttttaaatcctacatttgatgattga 959
Db 900 agactttagtgggaagtgtatatacaaaagtggttttaaatcctacatttgatgattga 959
Qy 960 ttgtcaacttcagatgatttcaatgaaatgataaacttctgggaacttaacatggcg 1019
Db 960 ttgtcaacttcagatgatttcaatgaaatgataaacttctgggaacttaacatggcg 1019
Qy 1020 cataaaactttggatgggtttcttactcgtatcgtctgcaaacatttggttcaaga 1079
Db 1020 cataaaactttggatgggtttcttactcgtatcgtctgcaaacatttggttcaaga 1079
Qy 1080 tccagatctctcgagacactttataatcaatgacaatgagtttcttcaactccaagatga 1139
Db 1080 tccagatctctcgagacactttataatcaatgacaatgagtttcttcaactccaagatga 1139
Qy 1140 ctgtgtacataattatctcccaaatcgaaaaaacacccgatgattacattctgtatacta 1199
Db 1140 ctgtgtacataattatctcccaaatcgaaaaaacacccgatgattacattctgtatacta 1199
Qy 1200 ccgaggtcgaaacgcgcattggatggatcggttgggtccgtgtatctacacccgaagccc 1259
Db 1200 ccgaggtcgaaacgcgcattggatggatcggttgggtccgtgtatctacacccgaagccc 1259
Qy 1260 gacactccccgaatcgatcattcccaaacctcaaaaagcagcccaaatccgtgggtcgaga 1319
Db 1260 gacactccccgaatcgatcattcccaaacctcaaaaagcagcccaaatccgtgggtcgaga 1319
Qy 1320 ctttaacaattttcatacaaacgcgaataatgttggggcctgagcctccatttgggtgaag 1379
Db 1320 ctttaacaattttcatacaaacgcgaataatgttggggcctgagcctccatttgggtgaag 1379
Qy 1380 gcttgagaaaaacacgcggaagggcgagaagtgtgtgataaaagaagctgtagagataga 1439
Db 1380 gcttgagaaaaacacgcggaagggcgagaagtgtgtgataaaagaagctgtagagataga 1439

```


Qy 1440 agaagaggttgaagagaggttgagaggttagagatactagatgaactttgtttcagag 1499
 |||||
 .Db 1440 agaagaggttgaagagaggttgagaggttagagatactagatgaactttgtttcagag 1499
 Qy 1500 gttgcttgaaggggtttaagaggttgcaacagatgaagagaattttgtgagggaggttgag 1559
 |||||
 Db 1500 gttgcttgaaggggtttaagaggttgcaacagatgaagagaattttgtgagggaggttgag 1559
 Qy 1560 taaagaagaagaagaaattctgaatgaacttcaaatggaaagcagactgaagttgaaaaagct 1619
 |||||
 Db 1560 taaagaagaagaagaaattctgaatgaacttcaaatggaaagcagactgaagttgaaaaagct 1619
 Qy 1620 ttttggcgcggttaccgattaggaacttagataaatttcgatgatattgattcagaaca 1679
 |||||
 Db 1620 ttttggcgcggttaccgattaggaacttagataaatttcgatgatattgattcagaaca 1679
 Qy 1680 tatatatagtcattgattgattttatttgcttgaatttgaatactttttgtatac 1739
 |||||
 Db 1680 tatatatagtcattgattgattttatttgcttgaatttgaatactttttgtatac 1739
 Qy 1740 gtataaacgtttgtgattttatttgcttgaatttgaatactttttgtatac 1799
 |||||
 Db 1740 gtataaacgtttgtgattttatttgcttgaatttgaatactttttgtatac 1799
 Qy 1800 ctttacaacaaattgattgtctattgttagccaagaggttcacgaaaagactgaaagg 1859
 |||||
 Db 1800 ctttacaacaaattgattgtctattgttagccaagaggttcacgaaaagactgaaagg 1859
 Qy 1860 tcttgcgggttgcgggttaggcacaaatttttggcggggagtcgcttgcgggttt 1919
 |||||
 Db 1860 ccttgcgggttgcgggttaggcacaaatttttggcggggagtcgcttgcgggttt 1919
 Qy 1920 tctttaaacaatgtatttttataaattgagttatttcaatttttgcgtaaaaaaa 1979
 |||||
 Db 1920 tctttaaacaatgtatttttataaattgagttatttcaatttttgcgtaaaaaaa 1979
 Qy 1980 aa 1981
 ||
 Db 1980 aa 1981

RESULT 2

AAT66242

ID AAT66242 standard; cdna; 1589 bp.

XX AC AAT66242;

XX DT 28-JUL-1997 (first entry)

XX DE Tobacco violaxanthin de-epoxidase cdna.

XX KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;

XX KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;

XX KW xanthophyll; lettuce; ss.

XX OS Nicotiana tabacum L. cv. xanthi.

XX FH Key Location/Qualifiers

XX FT CDS 42..1478

XX FT FT /*tag= a

XX FT transit_peptide 42..443

XX FT mat_peptide 444..1475

XX FT FT /*tag= c

XX PN W09717447-A2.

XX PD 15-MAY-1997.

XX PF 07-NOV-1996; 96WO-US18291.

XX PF 06-AUG-1996; 96US-0023502.

XX PR

PR 07-NOV-1995; 95US-0006315.

XX (CALJ) CALGENE INC.

XX PI Bugos RC, Rockholm DC, Yamamoto HY;

XX WPI: 1997-281036/25.

XX P-PSDB; AAW09875.

XX DNA encoding plant violaxanthin de-epoxidase - used to modify the

XX sensitivity of a plant to light

XX Disclosure; Fig 2; 4lpp; English.

XX A cdna clone (AAT66242) codes for tobacco violaxanthin de-epoxidase

XX (VDE) (AAW09875), an enzyme that catalyses the de-epoxidation of

XX violaxanthin to zeaxanthin and antheraxanthin. VDE nucleic acids

XX (see also AAT66241, AAT66243), in sense or antisense orientation, can

XX be used in genetic constructs, pref. also contg. a plastid

XX translocation sequence, to modify VDE levels in plants. Increased

XX levels result in the plant being tolerant of increased light and

XX therefore more productive and/or more resistant to disease.

XX Underexpression of VDE increases photosynthetic efficiency under

XX low light. The photosensitivity of a range of crops, trees and

XX ornamentals can be modified.

XX Sequence 1589 BP; 500 A; 261 C; 367 G; 461 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

QY	1083	agatcttctcgagcaactttataatcatgacaatgagtttcttoactacacaagatgactg	1142
Db	917	aagttatccgggatactctacaatcatgataatgagtatctctctacacaagatgactg	976
QY	1143	gtacatatattctcccaaatcgaacaacaccogtattacatatctgtactactacg	1202
Db	977	gtatatattgtcatccaaagttagaaaatagtcocagaggattacatatattgtctactataa	1036
QY	1203	aggctgaacacgacgatgggtatgatacoggttggttcogtggatctacacccgaagcccgac	1262
Db	1037	gggcagaaatgatcatgggatgggatacggctgtctgtactttacacagaagtgacgt	1096
QY	1263	actcccccgaatcgatctcccaaacctcaaaaagcagccaaaatccgtgggtcgcgactt	1322
Db	1097	tttgctgaaagcattataccggagttgcacaaccgcagctcaaaaagtgggcgtgattt	1156
QY	1323	taacaatttataacaacccgaacataagttgtggcctgagccctccattggtgaaaggct	1382
Db	1157	caacacattcataaaaacagacaaacactgtggccctgaaacctcccctttgtgagaggtt	1216
QY	1383	tgagaaaaacgcgcgaaggggcgagaagtgttgtataaaaagaagctgtagagatagaaga	1442
Db	1217	ggagagaagaagtggagaagggagaagcagcatcataaaagaagttgagagatagaaga	1276
QY	1443	agaggttgaaaaagaggtggagaaaggttagagatactgagatgactttgttcagaggtt	1502
Db	1277	agaagta-----gagaaggtgagagataaagaagtcacctatttcagtaaaact	1324
QY	1503	gcttgaagggtttaaggagttgcacaacagatgaagagaattttgtgaggaggttgagtaa	1562
Db	1325	gttgaaggtttaagagctccacgcagatgaagagaaactctctagagagcgtgagcaa	1384
QY	1563	agaagagaagaaattctgaatgaactcaaatgaaagcagctgaaattgtaaaagctttt	1622
Db	1385	agaagaaatgagttgttttggatggacttaaaatggaagcaactgaggtagaaaaactttt	1444
QY	1623	tgggcgcgcttaccgattagaaaaacttagataaaatttcgatgtattgatcagacaatat	1682
Db	1445	tgggcgtgccttaccacaagaagaaatgaaggcaagtatttttaaactatcaacatatat	1504
QY	1683	a 1683	
Db	1505	a 1505	

RESULT

AAT66243

ID AAT66243 standard; cDNA; 1555 BP.

XX

AC AAT66243;

XX
XX

DT 28-JUL-1997 (FIRST ENTRY)
VV

XX Arabidopsis violaxanthin de-epoxidase cDNA.

XX
CONFIDENTIAL

KW violaxanthin de-epoxidase; VDE; light; photosensitivity;

KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;

KW xanthophyll; lettuce; ss.

[illegible]

OS ARABIDOPSIS THALIANA CV. COLUMBIA.
yy

xx	Key	Location/Qualifiers
FH		

FT CDS 45..1433

ET /*tag= a

FT transit_peptide 45..383

```

FT      = 4 - 4.30
          / *tag= b
          384 1430

```

```

EPI mat_peptide
384.:1430
/*tag= C

```

XX
17

PN WO9717447-A2.

xix

PD 15-MAY-1997.

XX

PF	07-NOV-1996;	96WO-US18291.
XX		
PR	06-AUG-1996;	96US-0023502.
PR	07-NOV-1995;	95US-0006315.
XX		
PA	(CALJ) CALGENE INC.	
XX		
PI	Bugos RC, Rockholm DC, Yamamoto HY;	
XX		
DR	WPI; 1997-281036/25.	
DR	P-PSDB; AA009876.	
XX		
PT	DNA encoding plant violaxanthin de-epoxidase - used to modify the	
XX	sensitivity of a plant to light	
PT		
XX		
PS	Disclosure; Fig 3; 41pp; English.	
XX		
CC	A cDNA clone (AAT66243) codes for Arabidopsis violaxanthin	
CC	de-epoxidase (VDE) (AAW09876), an enzyme that catalyses the	
CC	de-epoxidation of violaxanthin to zeaxanthin and antheraxanthin.	
CC	VDE nucleic acids (see also AAT66241-42), in sense or antisense	
CC	orientation, can be used in genetic constructs, pref. also contg. a	
CC	plasmid translocation sequence, to modify VDE levels in plants..	
CC	Increased levels result in the plant being tolerant of increased	
CC	light and therefore more productive and/or more resistant to	
CC	disease. Underexpression of VDE increases photosynthetic	
CC	efficiency under low light. The photosensitivity of a range of	
CC	crops, trees and ornamentals can be modified.	
XX		
SO	Sequence 1555 BP; 485 A; 286 C; 375 G; 409 T; 0 other;	

Query Match	30.8%;	Score 610.2;	DB 18;	Length 1555;
Best Local Similarity	72.5%;	Pred. No. 2.2e-145;		
Matches 805; Conservative	0;	Mismatches 303;	Indels 3;	Gaps

Qy	550	ttatacaattggctatcgtattgtgtttgtcacatttgtatttcgtttccagagtgtatgcc	609
Db	324	ttgtctaaactcgtgggtgttttagctgcgcgtcccttatgtttccatctcgaatgcga	383
Qy	610	gttgatgctcttaaaacttgtcttgtttactcaaaagaatcaggattgagcttgcaaaa	669
Db	384	gttgatgctcttaaaacttgcattgttttgagggtgaggatagaaactcgcgaag	443
Qy	670	tgtatagcaaaacctcttgtgcgcaaacgttcctgtctacagacttgcacaactcgt	729
Db	444	tgaattgccacctgcctgtgcagccaatgcgcgtgccttcagacctgcataaacctg	503
Qy	730	ccttgacgagaccgaattgcagataaaatgtgtgacttgttcgaaaaacagtgtgtggac	789
Db	504	ccagatgaaccaggatgccagattaaatgtgggtatcgttttgagaaacagtgtgttgat	563
Qy	790	caattcaacagatgtgcggttttcccgaaagaataatgtgtcccggaatactggatgggt	849
Db	564	gagttcaacgagtgtctgtgtcgaaaaaaagtgtgttcctagaatacttgactcgcga	623
Qy	850	gaattcccggttccgcatcgaatgcagtgtgttcaaaattttaacatgaagacatttagt	909
Db	624	gaatttcctgcccagaccctctgttctgttacagaacttcaacatctcgactctaac	683
Qy	910	gggaagtgggtatatacaacagtggttttaactctacatttgatgacttgattgtcaactt	969
Db	684	gggaagtgggtacattacaagtggttgaaatccaacctttgatgccttcgaatgcagctg	743
Qy	970	catgagtttcatatggaaaatg---ataaacttgttgggaacttaacatggcgcataaaa	1026
Db	744	catgagttccacacagaaggtgcacaacaagcttgttggaaacatctcttggagaataaag	803
Qy	1027	acttggatgggtgttcttacttcgatctgtgtgcacaaacttgttcaagattccagat	1086
Db	804	accttagacagtggtgttttactaatgtgcagcgcacaaaaatctcgtgcagaatccctaac	863
Qy	1087	cttctctggagcaactttataatcatcatgacaatgagtttcttcactaccaagatgactggtac	1146

217 WWWWWW
218 WWWWWW

xx The present invention relates to a composition comprising two nucleic
cc acids each containing an electron-transfer group (ETM) having
cc different redox potentials. The invention is used for electronic
cc

XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2001, 23:48:41 ; Search time 166.68 seconds
(without alignments)
2691.702 Million cell updates/sec

Title: US-09-075-375A-1

Perfect score: 1981

Sequence: 1 tttgggttcgaattttacc.....aattttggcgtcaaaaaa 1981

Scoring table: IDENTITY_NUC

Gapex 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2.6/ptodata/2/ina/5A_COMB.seq:*

2: /cgn2.6/ptodata/2/ina/5B_COMB.seq:*

3: /cgn2.6/ptodata/2/ina/6A_COMB.seq:*

4: /cgn2.6/ptodata/2/ina/6B_COMB.seq:*

5: /cgn2.6/ptodata/2/ina/PCTUS_COMB.seq:*

6: /cgn2.6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	1981	100.0	1981	3	US-08-747-574-1
2	643.8	32.5	1589	3	US-08-747-574-2
3	611.8	30.9	1555	3	US-08-747-574-3
c 4	78	3.9	7218	1	US-08-232-463-14
c 5	44.8	2.3	51259	3	US-08-781-891-209
6	42.4	2.1	3211	2	US-08-574-959A-8
7	42.4	2.1	3211	4	US-09-357-014-8
8	42.4	2.1	3901	2	US-08-574-959A-6
9	42.4	2.1	3901	4	US-09-357-014-6
10	41.2	2.1	1505	1	US-07-915-246-1
11	41.2	2.1	9636	3	US-08-323-170B-1
c 12	39.6	2.0	16442	3	US-08-781-891-208
13	38.6	1.9	6942	2	US-08-460-309-3
14	38.6	1.9	6942	2	US-08-125-077-3
15	38.2	1.9	289	4	US-09-007-005-17
16	38.2	1.9	289	4	US-09-244-796-17
c 17	38	1.9	9048	3	US-08-973-273-4
c 18	38	1.9	24979	3	US-08-147-777-3
c 19	38	1.9	24979	3	US-08-452-872-3
c 20	38	1.9	24979	5	PCT-US93-03985-3
21	37.8	1.9	3891	1	US-08-480-604A-27
22	37.8	1.9	3891	2	US-08-405-496A-27
23	37.8	1.9	3891	4	US-08-915-136-27
c 24	37.4	1.9	6049	1	US-08-471-033-1
c 25	37.4	1.9	6049	2	US-08-471-044-1
c 26	37.4	1.9	6049	2	US-08-463-483A-1
c 27	37.4	1.9	6049	2	US-08-471-046A-1

c 28 37.4 1.9 6049 2 US-08-470-566B-1 Sequence 1, Appli

c 29 37.4 1.9 6049 2 US-08-469-334-1 Sequence 1, Appli

c 30 37.4 1.9 6049 3 US-09-300-529-1 Sequence 1, Appli

c 31 37.4 1.9 15378 3 US-08-785-420-1 Sequence 1, Appli

c 32 37 1.9 2030 2 US-08-705-937-7 Sequence 7, Appli

c 33 37 1.9 4982 3 US-08-699-103B-1 Sequence 1, Appli

c 34 37 1.9 5361 4 US-08-973-482-2 Sequence 2, Appli

c 35 37 1.9 6152 4 US-08-973-462-1 Sequence 1, Appli

c 36 37 1.9 53526 3 US-08-658-136-2 Sequence 2, Appli

c 37 37 1.9 53577 3 US-08-658-136-1 Sequence 1, Appli

c 38 36.8 1.9 3489 2 US-08-728-323A-1 Sequence 1, Appli

c 39 36.8 1.9 32207 2 US-08-770-379-20 Sequence 20, Appli

c 40 36.8 1.9 32207 4 US-08-757-669A-20 Sequence 20, Appli

c 41 36.6 1.8 8920 2 US-08-446-855A-1 Sequence 1, Appli

c 42 36.6 1.8 8920 4 US-09-150-741-1 Sequence 1, Appli

c 43 36.4 1.8 2085 1 US-08-072-070-1 Sequence 1, Appli

c 44 36.4 1.8 2085 1 US-08-465-746-1 Sequence 1, Appli

c 45 36.4 1.8 2085 1 US-08-214-164-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-747-574-1
; Sequence 1, Application US/08747574
; Patent No. 6015939
; GENERAL INFORMATION:
; APPLICANT: CALGENE, INC.
; TITLE OF INVENTION: PLANT VDE GENES AND
; TITLE OF INVENTION: METHODS RELATED THERETO
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 Inch, 1.40 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08747,574
; FILING DATE: No. 6015939ember 7, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/023,502
; FILING DATE: August 6, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/006,315
; FILING DATE: No. 6015939ember 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: 119-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1981
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
US-08-747-574-1

Query Match 100.0%; Score 1981; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1981; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tfggggttcgaatttaccaccacaaagtttgcctaccataaattgggataaagtgct 60
DB 1 TGGGGTTCGAATTTACCCACCACAAAGTTTGTCTACCAATAATTTGGATAAGGAGTCT 60

QY 61 aattcccttgtaacatttcccaatttctccctccgcacacacatatatactga 120
DB 61 AATTTCCCTTGTACAATTTTCCAAATTTCTCTCCGCCACACATATATATCTAGCG 120

QY 121 caattgaaagctacaatttggaaaaagacagagattttacaagacgagagataa 180
DB 121 CACTTGAACGCTTACAAATGTTTGAANAAGACGAGATTTTACAAGACGGAGAAATAA 180

QY 181 taagcttcaagtactccgactgcaggtggcctttggaagcccaacaaactggct 240
DB 181 TAAGCTTCAAGTACTCCGATCGTCAGTGGCTTGTGGAAGCCAAACAATGGCTATGGCT 240

QY 241 cttctcttcaactgtatttctctgcaagaggaagccctcaatttatatacaagatca 300
DB 241 CTTTCTCTTTCACACTGTATTTCTCTCAAGAGAGAGAGCCCTCAATTTATGCAAGATCA 300

QY 301 ccatgtaataagagtttccacagaggtggacaacctctaccacaataatcatgatgaaa 360
DB 301 CCATGTAATGAAGGTTTACACAGAGTGGACAACTCTTACCACAATAATCATGATGAAA 360

QY 361 attcgatccaaatggatatttatttcttctccggtgttttaactcttataagacaagt 420
DB 361 ATTCCGATCCAAACATGGATATTTTAAATTTCTTCCGGTGTGTACATCTTATAAGCAAGT 420

QY 421 tcttctcagattctagcattgcaaggtataatctcagatgcagatgcatacaagt 480
DB 421 TCYTTCACATTTCTAGCCATTCAGAGGTAATCTCAGATATGACATCGATACAAGT 480

QY 481 ttgaggaatacaaaagatttgatctcaaaagggggcatgactttgtattcttgaagaagcaa 540
DB 481 TTTGAGGAAATACAAAGATTTGTATCTCAAAAGGGGATGACTTTGTATCTTGAAGAGCAA 540

QY 541 tggagacaattcatacaattgctatcgattgttgcacattgttctcgtccaga 600
DB 541 TGGAGACAATTCATCAATTTGGCTATCGTATGTTGGTTCACATTTGTATCTGTTCCACA 600

QY 601 gttgatccgttgatgctctttaaacttgcgtttgttttaactcaaaagaatgcaggttag 660
DB 601 GTTGATCCGTTGATGCTCTTAAACTTTGCTGTGTTTACTCAAGAATGCAGGATTTGAG 660

QY 661 cttgcaaaatgatatagcaaaaccatcttgcgggcaaaagttgcctgtctacagactgc 720
DB 661 CTTGCAAAATGTATAGCAAAACCCATCTTGTGCGGCAAAAGCTTGCCTGTCTACAGACTTC 720

QY 721 acaatcgctcagcagaccgaatgcatagataaaatgtgtgactgttgcgaaacagt 780
DB 721 AACAAATCGTCTGACGAGACCGAATGTACATAAATGTGCTGACTTGTTCGAAACAGT 780

QY 781 gtggtgaccaattcaacagagtgcgggtttcccgaaagaataatgtgccccggaaatcg 840
DB 781 GTGCTGGAACCAATTAACAGAGTGTGCGGTTTCCGGAAGAAATGTGTGCCCCGGAATCG 840

QY 841 gatgtgggtgaattcccggttccggatcgtaatcgagtggttcaaaattttaacatgaaa 900
DB 841 GATGTGGGTGAATTTCCCGGTTCCGGATTCGTAATGCAAGTGTTCAAAATTTTAACATGAAA 900

QY 901 gactttagtgggaagtgtatatacaagtggttttaaatcctcatttgcattgtgat 960
DB 901 GACTTTAGTGGGAAGTGTATATAACAAGTGTGTTTAAATCCTACATTTGATGCAATTTGAT 960

QY 961 tgtcaacttcatgatttcatatggaataatgataaacttgttgggaacttcaacatggcgc 1020
DB 961 TGTCAACTTCAATGATTTTCAATGGAATAATGATAAATCTTGTGGAACTTTAAACATGGCGC 1020

QY 1021 ataaaaacttggatgggttcttcttactcgatctgctgcaaacatttgttcaagat 1080

RESULT 2

US-08-747-574-2

; Sequence 2, Application US/08747574

; Patent No. 6015939

; GENERAL INFORMATION:

DB 1021 ATAAAACTTTGGATGGTGTCTTTACTCGATCTGCTGCAAAACATTTGTTCAAGAT 1080
QY 1081 ccagatcttccctggagcaacttataatcatgacaatagagtttcttactacacagatgac 1140
DB 1081 CCAGATCTTCTCGAGCACATTTATAATCATGACAATGAGTTTCTTCTACCTACCAAGATGAC 1140
QY 1141 tgggtacatatattctcccaaatcgaaacacaaaccccgatgattacatatctgtatactac 1200
DB 1141 TGGTACATATATCTTCCCAATTCGAAATCGAAACAAACCCGATGATTACATATTCGTATACTAC 1200
QY 1201 cgaggtcgaaacgacgcacatgggatggatcggtgggtcccgatctacaccccgaaaccccg 1260
DB 1201 CGAGTCTGAAACGACGATGGATGGATGGATGGGTGGGTCCGTCTACACCCGAAACCCCG 1260
QY 1261 acactcccggaatgatcatcccaacctcaaaaagcagcaaaatcccggtgggtcgagac 1320
DB 1261 ACATCCCGCAATCGATCATCCCAACCTACAAAGACGACCAAAATCCGTGGGTTCGAGAC 1320
QY 1321 ttaacaatttcaatacaaacgcacaatagttgtgggctgagcctccattggtggaaagg 1380
DB 1321 TTTAAACAATTTTCAATAACAACCGACAATAGTTGTGGGCTCGAGCTCCATTTGGTGGAAAGG 1380
QY 1381 cttgagaaacagcgagagggcgagaggtgttgataaaaagaaagctgtagagatagaa 1440
DB 1381 CTTGAGAAACAGCGAGAGGGCGGAGAGTTGTTGATAAAAGAAAGCTGTAGAGATAGAA 1440
QY 1441 gaaaggttgaataaagaggttgagagaggttagagatactgagatgacttggtttcagagg 1500
DB 1441 GAAGAGGTTGAAAGAGGTTGAGAGGTTAGAGATCTAGAGATGACTTTGTTTCAGAGG 1500
QY 1501 ttgcttgaaggggttaaggaggttgcacacagatgaagagaaatttggagggaggttagt 1560
DB 1501 TTGCTTGAAGGGTTTAAAGGAGTTTCAACAAGATGAAGAGAAATTTTGTGAGGGAGTTGAGT 1560
QY 1561 aagaagaagaagaaatctgaatgaacttcaaatgaaagcgactgaagtggaaagctt 1620
DB 1561 AAAGAGAGAGAGAAATCTGAATGAACCTTCAATGGAAGCGACTGAAGTTGAAAGGCTT 1620
QY 1621 ttggggcgcgcttacogattaggaacttagataaaatttgcgatgattgattcagacaat 1680
DB 1621 TTTGGGCGCGGTTTACCGATTAGGAACTTAGATAAAATTTTCGATGATTGATTCAGACAAT 1680
QY 1681 atatactcataatgattatgatactagatactagaaaccccaaaaaacttttgtatagc 1740
DB 1681 ATATATAGTCATATGATGATATGATAGTACTAGAGAAACCCCAAAAAACATTTTGTATAG 1740
QY 1741 tgataaacqgtgtgtgattgtttgtttattggcttaaaattgtagaatagctttttaattc 1800
DB 1741 TGATAAACGTTTGTGTGATTTGTTTATTGGCTTAAATTTGTAATAGCTTTTAAATTC 1800
QY 1801 tttaaaaaaattgattgtctattgttagccaagaggttccacgaaaaagactgaaaggtt 1860
DB 1801 TTTACAAAAAATTTGATTGTCTATTGTGTAGCCAAGAGGTTCCAGAAAAAGACTGAAGAGGT 1860
QY 1861 cttgcggttgcgggttagccaaatttttggggcgggatcggtcttgcacgggtttt 1920
DB 1861 CTTCGGGTTTTCGGGTTTAGGCCAAATTTTGGCGCGGATCGTCTGTATCGGTTTT 1920
QY 1921 tctttaaacatgtatttttataaatgatgattatttcaatttttgggtctaaaaaaa 1980
DB 1921 TCTTTAAACATGTTATTTTATAAATGATGATTATTTTCAATTTTGGCTAAAAAAA 1980
QY 1981 a 1981
DB 1981 A 1981

; APPLICANT: CALGENE, INC.
 ; TITLE OF INVENTION: PLANT VDE GENES AND
 ; METHODS RELATED THERE TO
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Calgene, Inc.
 ; STREET: 1920 Fifth Street
 ; CITY: Davis
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 95616
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: Macintosh 7.5
 ; SOFTWARE: Microsoft Word 5.1a
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: 60/023,502
 ; FILING DATE: August 6, 1996
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: 60/006,315
 ; FILING DATE: No. 6015939ember 7, 1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Donna E. Scherer
 ; REGISTRATION NUMBER: 34,719
 ; NAME: Carl J. Schwedler
 ; REGISTRATION NUMBER: 36,924
 ; REFERENCE/DOCKET NUMBER: 119-2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (916) 753-6313
 ; TELEFAX: (916) 753-1510
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1589
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA to mRNA
 ; US-08-747-574-2

Query Match 32.5%; Score 643.8; DB 3; Length 1589;
 Best Local Similarity 71.8%; Pred. No. 4.1e-157;
 Matches 862; Conservative 0; Mismatches 327; Indels 12; Gaps 1;

QY	483	tgaggaaatacaagaatttgatctcctcaaaaggaggcgatgactttgattcttgaaagcaatg	542
DB	317	TAAGGGAAATTCAGGATGCAAAATTTCCAAAGATGTAGCTTTGATGGTTGGGAAATG	376
QY	543	gagacaattcatacaattggctatgattgttgacatttggatttcccgaggt	602
DB	377	GGCCCAATTTGCCAAACAGCAATTTAGCTATATTCATTTTGTGCTGCTTCAAAAGC	436
QY	603	tgatgcgttgatgctcttaaaactgtgttcttactcaagaagaatcagagattgaact	662
DB	437	TGATGCGGTGATGCTCTCAAGACTTGTACTTGTCTACTGAAAGAGTGCAGGTTAGAGCT	496
QY	663	tgcataatgatagcaaacccctcttctgctgcaaacgcttgctctcacagacttgcaa	722
DB	497	TGCGAAGTGCATTTCCAGCCCTGCATGTGCAGCTAATGTTGCTGCTCCAGACTTGCAA	556
QY	723	caatcctctgacgagacccaatgtcagataaaatgtggtgacttggatgcaaaacagttg	782
DB	557	CAATGACCTGACGAAACCGAATGTGAGATAAAATGTGCTGATTTGTTGAAACAGTGT	616
QY	783	ggtggaccaattcaacagagtgctgaggttcccgaaagaatgtgcccggaaatcgga	842
DB	617	CGTAGAGAGTTCATGAGTGTGCACTCTCCCGAAAGAAATGTGTACCTCGTAAATCTGA	676

QY	843	tgagggtgaattcccggttccggatcgtaataagtcagtggttcaaaattttaacatgaaaga	902
DB	677	TGTTGGTGACTTTCTGTACTGTATCCAGTGTCTTGTCCAGAAAGTTTGACATGAAAGA	736
QY	903	ctttagtggaagtgttatatacaagaagtggttttaactcctacatttgatgattgattg	962
DB	737	TTTTAGCGGGAATGGTTTCATTTACTCTGCGGGTTTGATCCCACTTTTGATGCTTTGATG	796
QY	963	tcaacttcatgagtttcatatgaaatgataaaacttgttgggaacttaacatgagcgcat	1022
DB	797	CCAATTGCATGAGTTCCATACAGAAAGAAACAAACTTGTGGGAATTTATCTTGAGAAAT	856
QY	1023	aaaaacttggatggtgtgttttcttactcgatctgctgtgaaacatttggttcaagatcc	1082
DB	857	ACGTACACTGTGATGAGGATTTTACTCGATCAGCGGTGCAAAATTCGTGCAAGATCC	916
QY	1083	agatcttccctggagcactttataatcatgacaatgagtttcttcaactccaagatgactg	1142
DB	917	AAAGTATCCGGGATACCTACAAATCATGATTAATGAGTATCTTCTCTACCAAGATGACTG	976
QY	1143	gtacataattatttcccaaatogaaaaaacaccccgatgattacatatctgtatactaccg	1202
DB	977	GTATATTTGTATCCAAAGTAGAAATAGTCCAGAGGATTACATATTTGTGTACTATAA	1036
QY	1203	aggtcgaaacacgcatggatgatacgggtgggtccctgctgatacccccgaagcccgac	1262
DB	1037	GGGCAAAATGATCATGGGATGGATAGTGGTGTCTGTCTACTTTACACAAGAGTGCACT	1096
QY	1263	actccccgaatcgatcatcccaaacctacaaaagcagcccaatccctgggtcgagactt	1322
DB	1097	TTTGCTGAAAGCAITATATACCGGAGTTGCAAAACCCAGCTCAAAAAGTTGGGCGTGATTT	1156
QY	1323	taacaatttatacaaacacccgacaaatagttgtggcctgagcctccatttgggtgaaagct	1382
DB	1157	CAACACATTCATAAAACAGACAATACATGTGGCCTGAACCTCCCTTGTGTGAGAGTT	1216
QY	1383	tgagaaaaacagcgagagggcgagaaagtgtgtgataaaaagaaactgtgagatagaaga	1442
DB	1217	GGAGAAAGTGGAAAGAGAGAAAGGACGATCATAAAAGTTGAGGAGATGAGAAGA	1276
QY	1443	agaggttgaataagaggtggagaaggttagagatactgagatgacttgttttcagaggtt	1502
DB	1277	AGAAGTA-----GAGAAGGTGAGAGATAAAGAAAGTACCCTTATTTCAGTAAACT	1324
QY	1503	gcttgaagggtttaaggaggttgcaacaagatgaagagaatttttgagggagttgagtaa	1562
DB	1325	GTTTGAAGGTTTAAAGAGCTCCAAACGAGATGAAGAGAACTTCTTAAGAGAGCTGAGCAA	1384
QY	1563	agaagagaagaaattcttgaatgaacttcaaatgaaagcgactgaagtgaagagctttt	1622
DB	1385	AGAAGAAATGGATGTTTGGATGGACTTTAAATGGAGCAACTGAGGTAGAAAACCTTTT	1444
QY	1623	tggcgcgcttaccgattaggaacttagataaatttcgatgattgattcagacaatat	1682
DB	1445	TGGCGTCTTACCAATAAGGAAATTAAGGTAAGTATTTTAAAAACTATCAACATATAT	1504
QY	1683	a 1683	
DB	1505	A 1505	

RESULT 3
 US-08-747-574-3
 ; Sequence 3, Application US/08747574
 ; Patent No. 6015939
 ; GENERAL INFORMATION:
 ; APPLICANT: CALGENE, INC.
 ; TITLE OF INVENTION: PLANT VDE GENES AND
 ; METHODS RELATED THERE TO
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Calgene, Inc.
 ; STREET: 1920 Fifth Street


```

; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/472,319
; FILING DATE: 30-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/919,951
; FILING DATE: 27-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9721
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6942 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-125-077-3

Query Match      1.9%; Score 38.6; DB 2; Length 6942;
Best Local Similarity 50.8%; Pred. No. 1.4;
Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 1373 tggaaagcgttgagaaacagcagagagggcgagaggtgttgataaaagagcgtgtag 1432
Db 5190 TGGAAAGGCTTCAGAAAGAGATTGACCAGATGATTAAAGAACTGAGGAGGAAAAATCTAG 5249

QY 1433 agatagaagaagcgttgaaagaggtggagaggttagagatactgagatgactttgt 1492
Db 5250 AGACACAAAGGAATTCGTCAGATGAGTTGGTAGCTGCAGAGGCCCTTCTCAAAAAAG 5309

QY 1493 ttcagaggttgctgaagggtttaagaggttgcaacaagatgaagagaattttgtgaggg 1552
Db 5310 TGAAGAAGCTGTTTGGAGAGTCCCGGGGGAAAAATGAAGAAATGGAGAGGATCTCCGGG 5369

QY 1553 a 1553
Db 5370 A 5370

RESULT 15
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G

```

US-09-007-005-17

```

Query Match      1.9%; Score 38.2; DB 4; Length 289;
Best Local Similarity 6.7%; Pred. No. 0.55;
Matches 15; Conservative 97; Mismatches 113; Indels 0; Gaps 0;

QY 1392 agcgaagagggcgagaaagttgttgataaaagacgtgtagagataaagaggttga 1451
Db 46 arcraarurgrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 105

QY 1452 aaaagaggtggagaaaggttagagatactgagatgactttgttcagaggttgc 1511
Db 106 nrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 165

QY 1512 gtttaagaggttgcaacaagatgaagagaattttgtgagggagttgagtaaaagagaa 1571
Db 166 nrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 225

QY 1572 ggaattctgaaatgaacttcaaatggaagcgactgaagtga 1616
Db 226 crugrcrgrurarcrcrururgrcrgcruaaaaaaaaaa 270

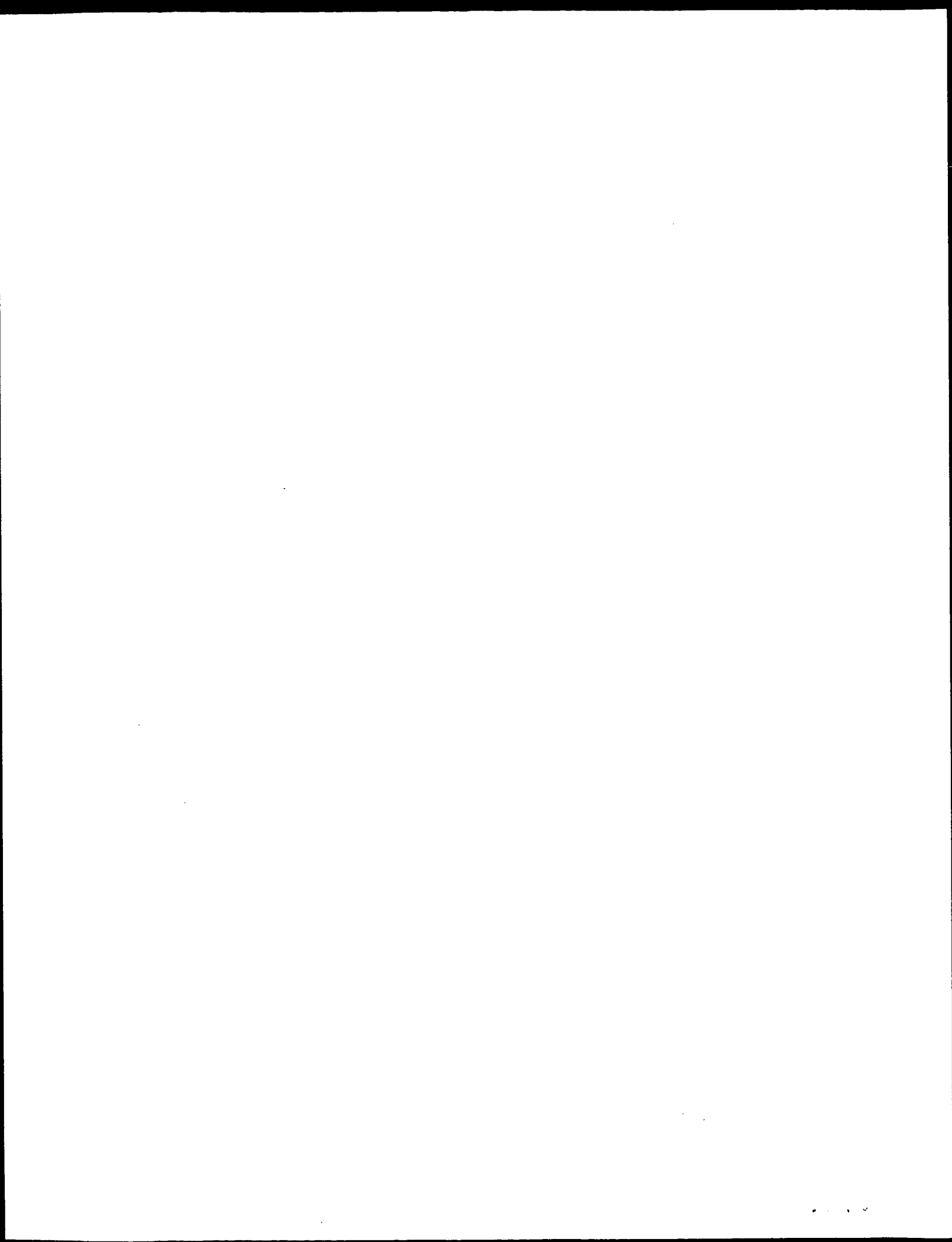
```

Search completed: November 6, 2001, 02:51:13
Job time: 10952 sec

Wed Nov 7 09:23:51 2001

us-09-075-375a-1.rni

Page 11



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2001, 23:45:51 ; Search time 3841.92 Seconds
(without alignments)
4874.152 Million cell updates/sec

Title: US-09-075-375A-1

Perfect score: 1981

Sequence: 1 tgtgggttcgaattttacc.....aattttggctaaaaaaa 1981

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	254.6	12.9	427	30	AV410579	AV410579 AV410579
2	249.2	12.6	483	104	AI994713	AI994713 701499353
3	240.8	12.2	515	144	BF069291	BF069291 st45b12.y
4	216.4	10.9	400	30	AV440942	AV440942 AV440942
5	181.2	9.1	438	159	N37612	N37612 18839 Lambd
6	175.8	8.9	248	30	AV410854	AV410854 AV410854
7	157.4	7.9	647	115	AW398615	AW398615 EST309115
8	154.2	7.8	871	234	AO858812	AO858812 nbab0002N
9	149.8	7.6	812	238	A2126249	A2126249 OSUNB0007
10	149.2	7.5	411	238	A2126934	A2126934 OSUNB0007
11	147.4	7.4	319	164	BE191198	BE191198 sn93d05.y
12	145.2	7.3	602	23	AI668224	AI668224 605018D09
13	143.2	7.2	599	256	B28051	B28051 T27G7TFB TA
14	141.2	7.1	335	30	AV521985	AV521985 AV521985
15	139	7.0	1147	256	B08217	B08217 T27M8-T7.1
16	137.8	7.0	724	174	BG123331	BG123331 EST468977
17	136.8	6.9	1147	256	B13003	B13003 T27G7-T7.1
18	135.6	6.8	549	141	BE918838	BE918838 FM1_2.F01
19	127.6	6.4	935	256	B12210	B12210 T27G7-T7 TA
20	127	6.4	985	256	B13009	B13009 T27M8-T7 TA
21	106.6	5.4	383	167	BE440946	BE440946 sp30a05.y
22	102	5.1	584	141	BE918982	BE918982 FM1_2.F01
23	97.8	4.9	483	151	BF623692	BF623692 HVSMEa000
24	96	4.8	418	137	BE595698	BE595698 P11_55.E1
25	89.6	4.5	584	20	AI483033	AI483033 EST242356
26	86.8	4.4	572	122	AW929268	AW929268 EST338056
27	81.6	4.1	112	174	BG157169	BG157169 sab23304.
28	74.8	3.8	1101	219	CNS0039G	AL063921 Drosophil
29	70.8	3.6	1101	219	CNS0039L	AL063926 Drosophil
30	66.8	3.4	1101	219	CNS0106X	AL098595 Drosophil
31	65.6	3.3	669	151	BF641377	BF641377 NF061G02I
32	62.8	3.2	1101	219	CNS0182P	AL108811 Drosophil
33	62.6	3.2	938	219	CNS006TJ	AL065906 Drosophil
34	62.6	3.2	1201	219	CNS0107R	AL098625 Drosophil
35	62	3.1	997	219	CNS005TE	AL060767 Drosophil
36	59.4	3.0	1101	219	CNS0039G	AL063921 Drosophil
37	58.6	3.0	928	219	CNS00DKY	AL071865 Drosophil
38	57.4	2.9	1086	219	CNS00YXK	AL096962 Drosophil
39	57	2.9	987	219	CNS00418	AL066537 Drosophil
40	56.8	2.9	1101	219	CNS00FVE	AL071298 Drosophil
41	56.6	2.9	967	219	CNS005Y0	AL062650 Drosophil
42	56	2.8	1201	219	CNS0169D	AL106459 Drosophil
43	55.8	2.8	1146	220	CNS021G2	AL176843 Tetraodon
44	55.4	2.8	1101	219	CNS00LT2	AL078714 Drosophil
45	55.2	2.8	1101	219	CNS008WC	AL052719 Drosophil

ALIGNMENTS

RESULT	1
AV410579	
LOCUS	AV410579 427 bp mRNA EST 23-MAY-2000
DEFINITION	AV410579 Lotus japonicus young plants (two-week old) Lotus
ACCESSION	Japonicus CDNA clone MWL074g05_r 5', mRNA sequence.
VERSION	AV410579.1 GI:7723433
KEYWORDS	EST.
SOURCE	Lotus japonicus.
ORGANISM	Lotus japonicus
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus.
AUTHORS	Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
TITLE	Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus

JOURNAL MEDLINE COMMENT

DNA Res. 7 (2), 127-130 (2000)

20277479

Contact: Yasukazu Nakamura
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source
Location/Qualifiers
1..427

/organism="Lotus japonicus"
/db_xref="taxon:34305"
/clone_lib="MWL074g05_r"
/dev_stage="young plants (two-week old)"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; isolate-MiyakoJima MG-20"
XhoI 112 a 85 c 102 g 128 t

BASE ORIGIN

Query Match 12.9%; Score 254.6; DB 30; Length 427;
Best Local Similarity 77.6%; Pred. No. 1e-51;
Matches 308; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 551 tcatacaattgctatcgtattggtttgcaattttgttatcttccagagttgatgcg 610
DB 31 TAATGAAGTGGCGGCATATTGGTGCATTTTAAATGATCGTCCACAGCTGATGCTG 90
QY 611 ttgatctcttaaaacttctgttttactcaagaatgcagagttgagcttgcacaaat 670
DB 91 TTGATGCTCTCAAACTTCTTCTTGTCTGAAGAAATGAGAGTGAAGTGGTGAATG 150
QY 671 gtatgcaaacccattctgttgcggcaaacgttgcctgtctacagacttgcaacaatcgtc 730
DB 151 GTATTCAAACCCGCTTGTGTGTCGCAATGTTGCTCTCCAAACTTGCAACAATCGAC 210
QY 731 ctgacgagaccgaatgcagataaaatggtgacttttcgaaacagttggtggacc 790
DB 211 CTGATGAGACCAATGCCAAATTAATGCGGGGACCTGTTTGAAGACAGTGGTGGTATC 270
QY 791 aattcaacgagttgctggtttccgcaagaatgtgtgccccggaatcgggtggtggtg 850
DB 271 AATTTAATGAGTGTGAGTCTCCCGGAAGAAATGTACTTAAGAAATCTGACGTGGGAG 330
QY 851 aattccgggttcgcatcgtatgagtcagtggttcaaaattttaacatgaagaacttagtg 910
DB 331 AGTTTCTCTCCAAATCCTGATGCTTGTGAAGAGCTTCAATATGCGAGATTTTCACTG 390
QY 911 ggaagtgttatatacaaatggttttaaatccctacatt 947
DB 391 GCAAGTGGTTCATCCTACTAGTGGCTTAATCTTACCTT 427

RESULT 2

AI994713
LOCUS AI994713 483 bp mRNA EST 08-SEP-1999
DEFINITION 701499353 A. thaliana, Ohio State clone set Arabidopsis thaliana
ACCESSION AI994713
VERSION AI994713.1 GI:5841618
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 483)
AUTHORS Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D., Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastyur, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobrifa, A., Murry, L., Turner, C., Krikorian, S., Elder, L. and Hanson, D.

Qy 603 tgatgcgttgatgctcttaaaactgtgctgtttactcaaaagaatgcagattgagct 662
 Db 464 TGATGCCGTGATGCTCTCAAACTTGTACTTGTCTACTGAAAGATGCGAGATAGACT 523
 Qy 663 tgcaaaatgtatgcaaacccattctgtgqgcaaacgttgcctgtctacagacttgcaa 722
 Db 524 TGCAGAGTGCATCTCAAACTGATGTCAGCTAAATGTTGCTCTCTCCAAACATGCRA 593
 Qy 723 caatgcctgcagagaccgaatgcagataaaatgtggtgacttcttgcgaacacagtg 782
 Db 584 CAATAGACCTGATGAAACGGAATGTGAGATAAAATGTGATTTGTTGAAAAACAGTGT 643
 Qy 783 ggt 785
 Db 644 TGT 646

RESULT 8
 A0858812/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

A0858812 871 bp DNA GSS 03-NOV-1999
 nbe0002N20r; CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
 clone nbe0002N20r, DNA sequence.
 A0858812
 A0858812.1 GI:6209269
 GSS.
 Oryza sativa.
 Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 871)
 Wing, R.A. and Dean, R.A.
 A BAC End Sequencing Framework to Sequence the Rice Genome
 Unpublished (1998)
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: GGAAACAGCTATGACCATG
 Class: BAC ends
 High quality sequence start: 27
 High quality sequence stop: 503.
 Location/Qualifiers
 1..871
 /organism="Oryza sativa"
 /lab_host="E. coli DH10B"
 /note="Vector: pBACindigo; Site.1: EcoRI; Site.2: EcoRI;
 Rice is the most important food crop in the world. Half of
 the world population, especially those inhabiting highly
 populated areas of the humid tropics and subtropics, rely
 on rice as their primary source of carbohydrate.
 Monocotyledonous rice is a diploid plant (2n=24) with a
 haploid genome equivalent of 431 Mbp (Arumuganathan and
 Earle, 1991). The relatively small genome of rice, three
 times larger than that of Arabidopsis, makes it suitable
 for genomic studies. In order to facilitate positional
 cloning, physical mapping and genome sequencing of rice,
 we have constructed a BAC library from Oryza sativa,
 Nipponbare variety using EcoRI as the cloning enzyme. The
 library contains 55,296 clones with an average insert size
 of 121 Kb providing approximately 15 haploid genome
 equivalents. The deep coverage allows the isolation a
 particular sequence with a probability of 99.9 %. Three
 high density filters, each containing 18,432 clones

(doubly spotted), represent the whole library for colony
 screening and can be requested from the Clemson University
 BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 243 a 181 c 197 g 245 t 5 others
 ORIGIN

Query Match 7.8%; Score 154.2; DB 234; Length 871;
 Best Local Similarity 69.3%; Pred. NO. 4.1e-27;
 Matches 210; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 744 atgtcagataaaatgtggtgacttctgcgaacacagtggtggaacacattcaacagatg 803
 Db 347 ATTGTAGATCAAAATGTGGAGATCTTTTGTGAGAACACTGTGTCGATGATGTTCAACGAGTG 288
 Qy 804 tgcggtttcccgaaagaaatgtgccccgggaaatcgagatggtggtgaattcccggttcc 863
 Db 287 TGCTGTTTCGCGCAAGAATGCGTCCACAAAAGTCCGACGTTGCGAGTTCACAGTCCC 228
 Qy 864 ggtatgtaatgcagtggttcaaaatttttaacatgaaagacttagtggaagtgttatat 923
 Db 227 TGATCCATCCGCCCTTGTCAAGAACTTCAACATGGCTGATTTCAACGGCAAGTGTATAT 168
 Qy 924 aacaagtgttttaaatcctacattggtgacattggtgacattgacattgacattgacatt 983
 Db 167 TTCAAGTGGCTCAATCCCACTTTTCGACACATTCGATTCGCAACTTCACGAGTTCGCGGT 108
 Qy 984 ggaatgataaaactgttgggaacttaacatggtggtggtggtggtggtggtggtggtt 1043
 Db 107 CGAGAGACAACTTATAGCAACTTCACATGGAGAATTCGAGCTCGTACCGGGGAT 48
 Qy 1044 ctt 1046
 Db 47 CCT 45

RESULT 9
 A2126249/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

A2126249 812 bp DNA GSS 02-JUN-2000
 OSJNB0075L17f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
 clone OSJNB0075L17f, DNA sequence.
 A2126249
 A2126249.1 GI:8200598
 GSS.
 Oryza sativa.
 Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 812)
 Wing, R.A. and Dean, R.A.
 A BAC End Sequencing Framework to Sequence the Rice Genome
 Unpublished (1998)
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: GTAAACGAGCGGCAGTG
 Class: BAC ends
 High quality sequence start: 40
 High quality sequence stop: 642.
 Location/Qualifiers
 1..812
 /organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="OSJNB0075L17f"
 /tissue_type="Leaf"

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

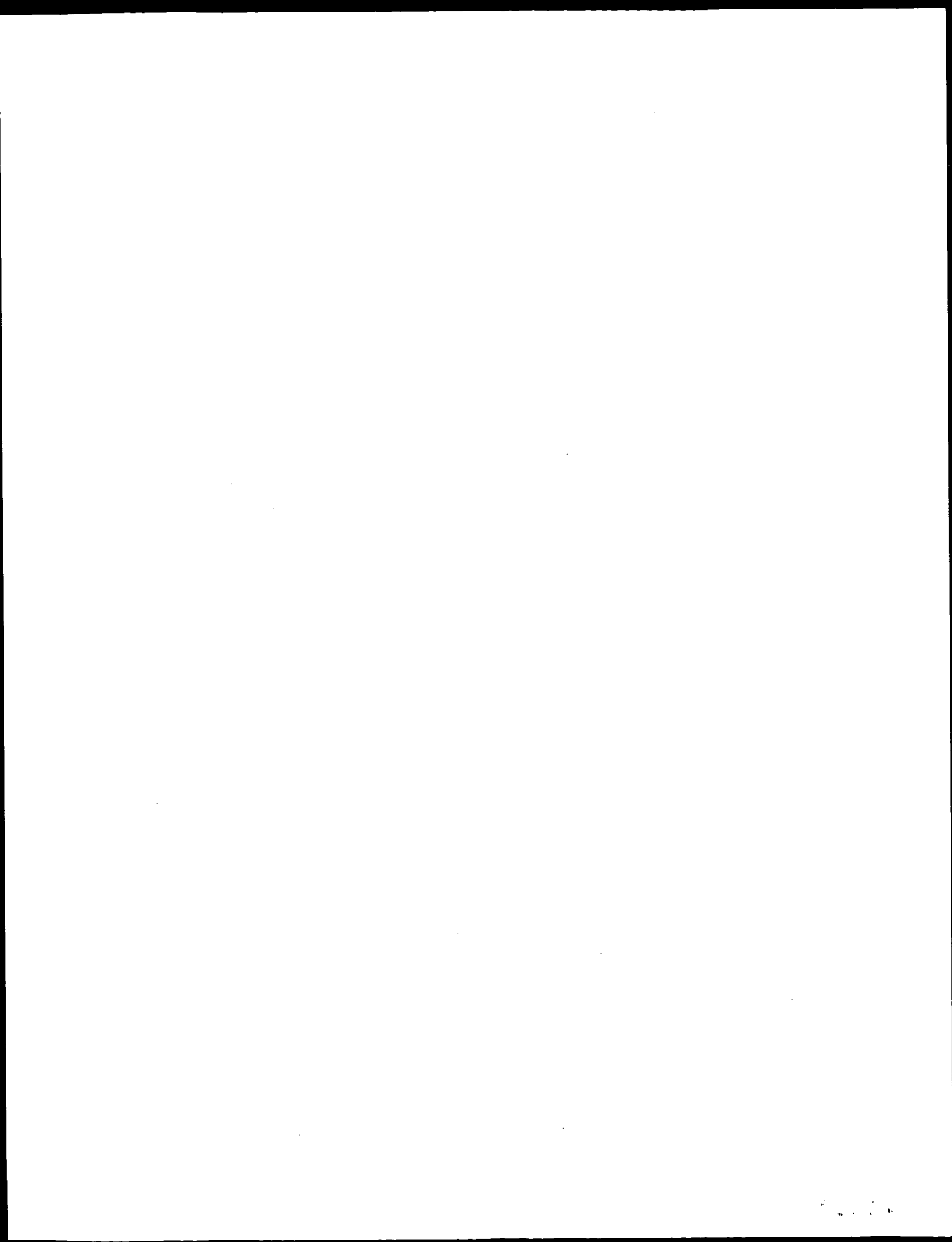
FEATURES
 source

/lab_host="E. coli DH10B"
/notes="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from *Oryza sativa*,
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
234 a 161 c 182 g 227 t 8 others
234 a 161 c 182 g 227 t 8 others

BASE COUNT		ORIGIN	
234	a	161	c
		182	g
		227	t
		8	others
Query Match 7.6%; Score 149.8; DB 238; Length 812;			
Best Local Similarity 69.0%; Pred. No. 4.8e-26;			
Matches 205: Conservative 0; Mismatches 92; Indels 0; Gaps 0;			

RESULT	10
AZ126934/C	
LOCUS	411 bp DNA GSS
DEFINITION	OSJNBb0079L12r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic clone OSJNBb0079L12r, DNA sequence.
ACCESSION	AZ126934
VERSION	AZ126934.1 GI:8202199
KEYWORDS	GSS.
SOURCE	Oryza sativa.
ORGANISM	Oryza sativa Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza. 1 (bases 1 to 411) Wing,R.A. and Dean,R.A. A BAC End Sequencing Framework to Sequence the Rice Genome Unpublished (1998)
REFERENCE	Contact: Wing RA
AUTHORS	Clemson University Genomics Institute
TITLE	Clemson University
JOURNAL	100 Jordan Hall, Clemson, SC 29634, USA
COMMENT	Tel: 864 656 7288

RESULT	11
BEI91198/c	
LOCUS	BEI91198
DEFINITION	sncgD05.yl 3m bp mRNA
ACCESION	BEI91198
EST	21-NOV-2000
Glycine max cDNA clone GENOME SYSTEMS CLONE ID:	
sn3Gd05.yl similar to TR:Q40251 VIOLAXANTHIN	
c1038-2146 5' similar to TR:Q40251 VIOLAXANTHIN	
DE-EPOXIDASE PRECURSOR.	; mRNA sequence.



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 02:52:36 ; Search time 93.61 Seconds
(without alignments)
914.444 Million cell updates/sec

Title: US-09-075-375A-2

Perfect score: 7495

Sequence: 1 MALSLHTVFLCKEALNYA.....MEASEVEKLFKALPIRKVR 1412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SIDSL1/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2554	34.1	478	18 AAW09875	Tobacco violaxanthin
2	2498	33.3	473	18 AAW09874	Romaine lettuce vi
3	2417	32.2	462	18 AAW09876	Arabidopsis violax
4	212.5	2.8	525	21 AAG49996	Arabidopsis thalia
5	210	2.8	522	21 AAG49997	Arabidopsis thalia
6	209	2.8	522	21 AAG06331	Arabidopsis thalia
7	209	2.8	526	21 AAG06330	Arabidopsis thalia
8	207	2.8	428	21 AAG49998	Arabidopsis thalia
9	206	2.7	428	21 AAG06332	Arabidopsis thalia
10	151	2.0	1055	21 AAY44787	Arabidopsis thalia
11	147.5	2.0	1786	18 AAW24790	P. falciparum live

12	147	2.0	2482	16 AAR72826	Human mitotin. Ho
13	147	2.0	2482	19 AAW23996	Human mitotin amin
14	143.5	1.9	1558	21 AAB18324	Plasmodium falcipa
15	135.5	1.8	1312	18 AAW22775	Human RAD50. Homo
16	135.5	1.8	1312	19 AAW71295	Human homologue of
17	135	1.8	968	19 AAW85011	p8alpha-green flo
18	134	1.8	1639	19 AAW54145	P. falciparum synt
19	133	1.8	1979	21 AAB18171	Plasmodium falcipa
20	132	1.8	776	15 AAR60178	Lethal factor of B
21	129.5	1.7	2954	20 AAY01632	Amino acid sequenc
22	129	1.7	993	17 AAR95268	Pre-nisin modifica
23	129	1.7	993	20 AAY06666	Nisin B of lactoba
24	129	1.7	1654	6 AAP50777	Sequence of the PI
25	128.5	1.7	1521	21 AAG39235	Arabidopsis thalia
26	128.5	1.7	1528	21 AAG39234	Arabidopsis thalia
27	128.5	1.7	1562	21 AAG39233	Arabidopsis thalia
28	128.5	1.7	1703	21 AAG36714	Arabidopsis thalia
29	128.5	1.7	1710	21 AAG36713	Arabidopsis thalia
30	128.5	1.7	1744	21 AAG36712	Arabidopsis thalia
31	124	1.7	1498	16 AAR77084	Rat sulphonylurea
32	124	1.7	1582	16 AAR77087	Chicken leucocytos
33	121.5	1.6	1132	17 AAR97866	Human secreted pro
34	121.5	1.6	1369	20 AAY24788	Nucleolar/endosoma
35	121	1.6	1411	17 AAW02258	Mutant C-beta prot
36	119.5	1.6	1099	19 AAW40538	TNF-R-EBA 175 fusi
37	119.5	1.6	1604	16 AAR70105	M. jannaschii MJ14
38	119.5	1.6	1847	21 AAY52002	M. jannaschii MJ14
39	119.5	1.6	1847	21 AAY51631	Streptococcus pneu
40	119.5	1.6	1881	21 AAY44506	Human p160 polyep
41	119	1.6	905	18 AAW31186	Human p160 polyep
42	119	1.6	1135	18 AAW31185	Erysiphe graminis
43	118.5	1.6	2273	17 AAR98811	Potato starch synt
44	117.5	1.6	1230	21 AAB49306	P. falciparum FCR3
45	117.5	1.6	3542	22 AAB62142	

ALIGNMENTS

RESULT 1
AAW09875
ID AAW09875 standard; Protein; 478 AA.
XX
AC AAW09875;
XX
DT 28-JUL-1997 (first entry)
XX
Tobacco violaxanthin de-epoxidase.
XX
Violaxanthin de-epoxidase; VDE; light; photosensitivity;
KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
KW xanthophyll; tobacco.
XX
Nicotiana tabacum cv. xanthi.
XX
FH Key Location/Qualifiers
FT Peptide 1..134
FT /label= Transit_peptide 135..478
FT Protein /label= Mat_protein 135..147
FT Peptide /note= "Claim 8"
FT Domain 135..206
FT /label= Cys-rich_domain 385..478
FT Domain /label= Highly-charged_domain 141
FT Misc-difference 141 /note= "conserved Cys residue"
FT Misc-difference 143 /note= "conserved Cys residue"
FT Misc-difference 148 /note= "conserved Cys residue"
FT Misc-difference 155 /note= "conserved Cys residue"

FT Misc-difference 161 /note= "conserved Cys residue"
 FT Misc-difference 167 /note= "conserved Cys residue"
 FT Misc-difference 171 /note= "conserved Cys residue"
 FT Misc-difference 180 /note= "conserved Cys residue"
 FT Misc-difference 184 /note= "conserved Cys residue"
 FT Misc-difference 190 /note= "conserved Cys residue"
 FT Misc-difference 206 /note= "conserved Cys residue"
 FT Misc-difference 252 /note= "conserved Cys residue"
 FT Misc-difference 382 /note= "conserved Cys residue"

XX WO9717447-A2.

XX 15-MAY-1997.

XX 07-NOV-1996; 96WO-US18291.

XX 06-AUG-1996; 96US-0023502.

XX 07-NOV-1995; 95US-0006315.

XX (CALJ) CALGENE INC.

XX Bugos RC, Rockholm DC, Yamamoto HY;

XX WPI: 1997-281036/25.

XX N-PSDB: AAT66242.

XX DNA encoding plant violaxanthin de-epoxidase - used to modify the sensitivity of a plant to light

XX Disclosure; Fig 2; 41pp; English.

XX The 55 kDa violaxanthin de-epoxidase (VDE) (AAW09875) of tobacco catalyzes the de-epoxidation of violaxanthin to zeaxanthin and antheraxanthin. This system, termed energy dependent non-radiative energy dissipation or non-photochemical fluorescence quenching, reduces the quantum efficiency of photosystem II (PSII), helping to prevent PSII over-reduction and photoinhibitory damage. The amino acid sequence of the VDE was deduced from an isolated cDNA clone (AAT66242). VDE nucleic acids (see also AAT66241, AAT66243), in sense or antisense orientation, can be used in genetic constructs to modify VDE levels in plants. Increased levels result in the plant being tolerant of increased light and therefore more productive and/or more resistant to disease. Underexpression of VDE increases photosynthetic efficiency under low light. The photosensitivity of a range of crops, trees and ornamentals can be modified.

XX Sequence 478 AA;

Query Match 34.1%; Score 2554; DB 18; Length 478;

Best Local Similarity 99.8%; Pred. No. 3.9e-191;

Matches 476; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 474 MALAPHSFLANHETIKYVGSKLPGHRSFGWGHEDFGSIVVAKICSSRRIPRYFKSP 533

Db 1 malaphsflanhetikyyvgsklpgghrfswgwyfgsivvakiessrripryfksp 60

OY 534 RTCCGLDRLGQLFSGHKNLSPAHNSNONVPKNSGCKPKDVALMWKQWQFAKTAI 593

Db 61 riccglgrgvlqfshgkmlspahnsinqvpkngscgkfpkdvalmwkqgafaktai 120

OY 594 VAIFILSVASKADAVDALKTCTLLKRELRLEAKISNPACAAANVACLQTCNNRPDETEC 653

Db 1 vaifilsvaskadavdalktctllkrclelrleakispacaaanvacqltcnnrpdetec 171

Db 121 vaifilsvaskadavdalktctllkrclelrleakispacaaanvacqltcnnrpdetec 180
 OY 654 QIKGDLFENSVDDEFNECAVSRKKCVPRKSDVGDFPVPDPSVLVQKDFDKDFSGKWFIT 713
 Db 181 qkcgdlfensvvddefnecavsrkkcvprksdvgdfpvpdpsvlvqkdfdkdfsgkwfit 240
 OY 714 RGLNFTFADFDCQLHEFTEENKLVGNLSWIRTPDGGFFTRSAVQKFPVQDPKYPGILYN 773
 Db 241 rglntfdafdcqlhefhteenklvgnlswirtpdggfftrsavqkfvqgkypgilyn 300
 OY 774 HDNEVLLYQDDWYILSSKVENSPEDYIFVYKGRNDADWGGVSVLYTRSAVLPESTIPE 833
 Db 301 hdneyllyqddwyilsskvenspedyifvykgrndawdggvsvlytrsavipesipe 360
 OY 834 LQTAACKVGRDENTFIKTDNTCGPEPLVERLEKKVEEGERTIIKEVEEIPREEVEKVRDK 893
 Db 361 lqtaackvgrdnfntfiktntcgppeplverlekkveegertiikeveeipreevekvrk 420
 OY 894 EYTLFSKLFEFGKELQDRENFRLSKKEEMDVLGLKMEATEVEKLFGRALPIRKL 950
 Db 421 evtlfsklfegfkqlrdeenflrslskeemdvldglkmeateveklfgralpirkl 477

RESULT 2

AAW09874

ID AAW09874 standard; Protein; 473 AA.

AC AAW09874;

XX 28-JUL-1997 (first entry)

XX Romaine lettuce violaxanthin de-epoxidase.

XX Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 KW xanthophyll; lettuce.

XX Lactuca sativa L. cv. romaine.

Key	Location/Qualifiers
FT Peptide	1..125
FT Protein	/label= Transit_peptide
FT Peptide	126..473
FT Domain	/label= Mat_protein
FT Peptide	126..138
FT Domain	/note= "Claim 8"
FT Peptide	126..197
FT Domain	/label= Cys-rich_domain
FT Peptide	218..231
FT Domain	/label= Lipocalin_signature
FT Peptide	376..473
FT Domain	/label= Highly-charged_domain
FT Peptide	265..272
FT Peptide	/label= Tryptic_peptide-11
FT Peptide	275..289
FT Peptide	/label= Tryptic_peptide-21
FT Peptide	341..353
FT Misc-difference 132	/label= Tryptic_peptide-15
FT Misc-difference 134	/note= "conserved Cys residue"
FT Misc-difference 139	/note= "conserved Cys residue"
FT Misc-difference 146	/note= "conserved Cys residue"
FT Misc-difference 152	/note= "conserved Cys residue"
FT Misc-difference 158	/note= "conserved Cys residue"
FT Misc-difference 162	/note= "conserved Cys residue"
FT Misc-difference 171	/note= "conserved Cys residue"

FT Misc-difference 175 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 190 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 197 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 243 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 373 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 XX WO9717447-A2.
 XX 15-MAY-1997.
 XX
 PF 07-NOV-1996; 96WO-US18291.
 XX
 PR 06-AUG-1996; 96US-00233502.
 PR 07-NOV-1995; 95US-0006315.
 XX
 PA (CALJ) CALGENE INC.
 XX
 PI Bugos RC, Rockholm DC, Yamamoto HY;
 XX
 XX WPI; 1997-281036/25.
 DR N-PSDB; AAT66241.
 XX
 XX DNA encoding plant violaxanthin de-epoxidase - used to modify the
 FT sensitivity of a plant to light
 FT
 PS Example 1; Fig 1; 4lpp; English.
 XX
 CC The 55 kba violaxanthin de-epoxidase (VDE) (AAW09874) of romaine
 CC lettuce catalyses the de-epoxidation of violaxanthin to zeaxanthin
 CC and antheraxanthin. This system, termed energy dependent
 CC non-radiative energy dissipation or non-photochemical fluorescence
 CC quenching, reduces the quantum efficiency of photosystem II (PSII)
 CC helping to prevent PSII over-reduction and photoinhibitory damage.
 CC The amino acid sequence of the VDE was deduced from an isolated
 CC cDNA clone (AAT66241). VDE nucleic acids (see also AAT66242-43), in
 CC sense or antisense orientation, can be used in genetic constructs
 CC to modify VDE levels in plants. Increased levels result in the
 CC plant being tolerant of increased light and therefore more
 CC productive and/or more resistant to disease. Underexpression of
 CC VDE increases photosynthetic efficiency under low light. The
 CC photosensitivity of a range of crops, trees and ornamentals can be
 CC modified.
 XX
 XX Sequence 473 AA;
 SQ
 Query Match 33.3%; Score 2498; DB 18; Length 473;
 Best Local Similarity 99.6%; Pred. No. 9.2e-187;
 Matches 471; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MALSLHTVFLCKEALNLYARSPCNERFHRSGOPPTNIIMMKIRSNNGYNFSRLTYSK 60
 DB 1 malslhtvfickealnlyarspcnerfhrsgopptnlmmkirsnnngynfsrltysk 60
 QY 61 TSSFSDSSHCKDKSQICSIDTSFEIQRFDLKRGMWLILEKQWRQFIQLAIVLVCTFVIV 120
 DB 61 tssfsdsshckdksqicsidsfteelqrfdkrgmtlilekqrqfiqlaivlvctfiv 120
 QY 121 PRVDAVDAKTCACLLKECRIEIAKCIANPSCAANVACLQTCNNRPDETCQIKCGDLFE 180
 DB 121 prvdavdalktcacllkecrielakciianpscaanvACLqtcnnrpdetecqikcgdlfe 180
 QY 181 NSVVDQNECAVSRKCVPRKSDVGEPVPDRNAVQNFNMKDFSGKWTYISGLNPTFDA 240
 DB 181 nsvvdqnecavsrkvprksgdvgevpvdrnavqnfnmkdfsgkwyitsglnptfda 240
 QY 241 FDCQLHEFHMKNDKLVGNLTWRIKTLGGFFTRSAVQTFVQDDPLPCGALYNHNDNEFLHYQ 300

DB 241 fdcqlhefhmkndkvlgnltwriktldgfftrsvagtqvqdpdlpgalynhndneflhyq 300
 QY 301 DDWYILSSQIENKPDYIFVYGRNDADWDGYSVIYTRSPITPESIIPLNQLKAAKSVG 360
 DB 301 ddwylssqienkpdyytffygrndawdgygsviytrspitpestipnlqkaaksvg 360
 QY 361 RDNFNITTDNSCGPEPLVERLEKTAERGEKILLIKEAVEIEEEVEKEVKVDRDTEMTLF 420
 DB 361 rdnfnittndnscgpeplverlektaeegekillikeaveieeeevekekvrdtemtlf 420
 QY 421 QRLLEGFKELQDDENFVRELSKEKEILLNLOMEATEVEKLFGRALPIRKLR 473
 DB 421 qrllegfkellqdeenfvrelskeekellnelqmeateveklfgralpirkir 473
 XX
 RESULT 3
 AAW09876
 ID AAW09876 standard; Protein; 462 AA.
 XX
 AC AAW09876;
 XX
 DT 28-JUL-1997 (first entry)
 XX
 DE Arabidopsis violaxanthin de-epoxidase.
 XX
 KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 KW xanthophyll.
 XX
 OS Arabidopsis thaliana var. columbia.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..113
 FT /label= Transit_peptide
 FT Protein 114..462
 FT /label= Mat_protein
 FT Peptide 114..126
 FT /note= "Claim 8"
 FT Domain 114..185
 FT /label= Cys-rich_domain
 FT Domain 364..462
 FT /label= Highly-charged_domain
 FT Misc-difference 120 /note= "conserved Cys residue"
 FT Misc-difference 122 /note= "conserved Cys residue"
 FT Misc-difference 127 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 134 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 140 /note= "conserved Cys residue"
 FT Misc-difference 146 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 150 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 159 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 163 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 178 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 185 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 231 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 362 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 XX WO9717447-A2.
 XX 15-MAY-1997.
 PD

```
XX 07-NOV-1996; 96WO-US18291.
PF
XX 06-AUG-1996; 96US-0023502.
PR 07-NOV-1995; 95US-0006315.
XX
XX (CALJ ) CALGENE INC.
XX
XX Bugos RC, Rockholm DC, Yamamoto HY;
XX
XX WPI; 1997-281036/25.
DR N-PSDB; AAT66243.
XX
XX DNA encoding plant violaxanthin de-epoxidase - used to modify the
XX sensitivity of a plant to light
XX
XX Disclosure; Fig 3; 41pp; English.
XX
XX The violaxanthin de-epoxidase (VDE) (AAW09876) of Arabidopsis
XX catalyses the de-epoxidation of violaxanthin to zeaxanthin and
XX antheraxanthin. This system, termed energy dependent non-radiative
XX energy dissipation or non-photochemical fluorescence quenching,
XX reduces the quantum efficiency of photosystem II (PSII), helping to
XX prevent PSII over-reduction and photoinhibitory damage. The amino
XX acid sequence of the VDE was deduced from an isolated cDNA clone
XX (AAT66243). VDE nucleic acids (see also AAT66241-42), in sense or
XX antisense orientation, can be used in genetic constructs to modify
XX VDE levels in plants. Increased levels result in the plant being
XX tolerant of increased light and therefore more productive and/or
XX more resistant to disease. Underexpression of VDE increases
XX photosynthetic efficiency under low light. The photosensitivity of
XX a range of crops, trees and ornamentals can be modified.
XX
XX Sequence 462 AA;

Query Match 32.2%; Score 2417; DB 18; Length 462;
Best Local Similarity 99.8%; Pred. No. 1.9e-180;
Matches 461; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 951 MAVATHCTSPCHDRIRFFSDGIGRLGTRKTRNGTFLKLLPIQSAADLRTTGRRSS 1010
Db 1 mavathctspchdrirffsdgigrlgtrkrngtflkllppiqaadlrrtgrrss 60
QY 1011 RPLSAFRSGFSKGFIDVPLPSKNEKELPAPLLLLKLVGLACAFILVPSADAVDAKTC 1070
Db 61 rplsafrsgfskgifdlvplpsknelkeltapllllklvgvlacafilvpsadavdalktc 120
QY 1071 ACLKGCRIELAKCIANPACAANVACIQTCCNRPDETECOIKGDLFENSVDDEFNECAV 1130
Db 121 aclkgrielakcianpacaanvaciqtcnrrpdeceqikcgdlfensvvddefnecav 180
QY 1131 SRKCVPRKSDGLGFAPPDPSVLVQNFNISDFNGKWYITSGLNPTDFADQQLHEFHTEG 1190
Db 181 srkcvprksdglgfappdpssvlvqnfnsdfngkwytsglnptdfadqqlhefhteg 240
QY 1191 DNKLVGNISRIKTLDSGFTTRSAVQKVFQDPNPGVLYNHNDNEYLHYQDDWYILSSKIE 1250
Db 241 dnklvgnlisrliktdsgfttrsavqkvfqpdpnpgvlynhndneylhyqddwylsskie 300
QY 1251 NKPEDYIFVYVRGRNDAMDCYGGAVVYTRSSVLPNSTIPELEKAASIGRDFSTFIRTDN 1310
Db 301 nkpedyifvyvrgrndacdyygavvytrssvlpnslipelekaasigrdfstfirtdn 360
QY 1311 TCGPEPALVERIEKTVEGEIRIIVKEVEETEEVEKEVEKVGRTMTLFORLAEGNELK 1370
Db 361 tcgpepalveriektveegeriivkeveeeteevekevekvgrtmtlforlaegfnelk 420
QY 1371 QDENFVRELSKREMEFLDETKMEASVEKLFKALPIRKVR 1412
Db 421 qdenfvrelskeemefldekmeaseveklfgkalpirkvr 462
```

RESULT 4
AAG49996
ID AAG49996 standard; Protein; 525 AA.
XX
AC AAG49996;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63310.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.


```
QY 447 EILNELQMEATEVEKLFGRALPIRKLRLMALPHSNFLANHETIKYVVGSKLPCHGRKFSWG 506
Db 128 ramn-ltqelnqtdilvvva-----vnasesv-----n 154
QY 507 WEDYFGSIVVAKICSSRRIPRYFRKSPRICCGLDRLQGLFSGKHNLSLPAHSINQNVPK 566
Db 155 wigtstknvnmic-----fessnlnmr1gg-----tdvgsnvkd--- 190
QY 567 GNSCKPKPKDALMV-----WKWG-----QFAKTAIValFILSV-----ASKADAVDA 610
Db 191 -----kvevkvgtvgdawerinsddirfcflvlinayirpvpqlnlsrk-----g 237
QY 611 LKTCGTLLEKCELELACISNPAACAAVACLQTCNNRPDETEQIKCGDLFNSVVDDEFN 670
Db 238 fstlscmvkncgqilnclldpncrkalcqlnqcs--pvdqvcysrciasyeypeafs 295
QY 671 ECAVSRKKCVPRKSDVGFPPDPDSVLVQKFDKMDPSGK----- 709
Db 296 lcvlqkhncleldakipekpyvvp-----mtsfgrkelchdtaedlfvgwlgelews 347
QY 710 WFTIRGLNPFDAFCQLHEFH-----TEENKLVGNLSWRIRTPDGGF 752
Db 348 wrvvagqpnydqfpcyqlfyrkgktsfwyepvfqrileeklv-----wrrr----- 397
QY 753 FTRSAVQKVFQDPKYP-----GILYNH--DNEYLLYQDDWYILSSKVENSPEDYIFVYK 805
Db 398 --rysvkr-----gkipatfrfsvldngvsnef-----wtivd--vsddlsvglfhyhg 443
QY 806 GRNDWDGCGYGSVLYTRSAVLP--RSIIPETQAAQKVG--RDNFTFIKTDNTCGPEPLVE 863
Db 444 aarvagsygtgavilvtpdgsypaekdkerlqsaalekcgikewelf-avdnscenppl-- 500
QY 864 RLEKKVEGER-----RTIIEVEIEE 885
Db 501 ----gipqgsrlhrislieepdseek 523

RESULT 5
AAG49997
ID AAG49997 standard; Protein; 522 AA.
XX AC AAG49997;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 63311.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
```


XX Arabidopsis thaliana protein fragment SEQ ID NO: 3067.
DE Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
KW
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138544.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145216.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.

PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 2.8%; Score 209; DB 21; Length 522;		
Best Local Similarity 21.2%; Pred. No. 8.4e-08;		
Matches 122; Conservative 76; Mismatches 191; Indels 186; Gaps		
QY	375	PEPPLVERLEKTAEGEKKLKEAVEIEEVEEKEVKEVTEMTLFORLLEGFKEQLQDE 434
Db	68	peppv--klialvgkgevsplk-stsweevmlhtarlkwwd-----egyemlvfdd 116
QY	435	ENFYRELSKEEKILNELQWEATEVEKLFGRALPIRLKLMALAPHSNFIANHETIKYYVG 494
Db	117	e-----llssndqtalt-t-lkgeinqtldilvvva-----vnnsevs----- 150
QY	495	SKLPGHKRFSGWEDYFGSIVVAKICSSRRIPRYFRKSPRICGLDSRGLQLFSGCHKNL 554
Db	151	-----nwigtstgnvknmic-----fesspnlmr1gg----- 178
QY	555	SPAHSINQNPVKGNSCKFPKDALMV-----WEKVG-----QFAKTAIVAIFILSV-- 601
Db	179	tdvgsvnkd-----kevtevkttvedawerrnsddirfc1lviinayirpvpvl 227
QY	602	---ASKADAVDAKTCCTLLKCEKRELELAKICISNPACAAVACIQTCTCNRPDETCQIKCG 658
Db	228	qn1rsk-----gfstlscmknccpqilnc1ldpncrkalc1qcInqcs--pvdqvcsvrci 280

QY	659	DLEPNSVVDENFECASVRKCKCVPKRSDVGDFPVDPDPSVLVQKFDMKDFSGK-----	709
		: : :	
Db	281	asyespyfeafslcvlqhklncldekdakipkvypvp-----mtsfrgkelchdtaed	332
QY	710	-----WFIITRGLNPTFDAFCQLHEPHTEENK-----LVGNLSWR	744
		: :	
Db	333	lfvgwlgelewsrvvragnpaydqfpcdyqlfyrgkgkssfwyepvfqvrtlegklivr	392
QY	745	IRTPDGGFETRSVAVQKFVODPKYP-----GILYNH--DNEYLLYQDDWILSSKVENSPE	797
		: :	
Db	393	rr-----rysvkr-----gkipatfrfsvidngvvsnef-----wtivd--vsddls	432
QY	798	DYIFVYVKGNDAWDGGSVLTVTRSAVIP-ESIIPLOTAAQKVG-RDNFTIKTDNTC	855
		: : :	
Db	433	wglfhynagarvagqgytgvilvtpdgspaekekerlgksalekcgcikewelf-avdnccs	491
QY	856	GPPPIPLVERLEKKVEEGER-----TIKEVEEEIEE	885
		: :	
Db	492	cenppl-----gipqgsrlhsrisieeepdseek	520
RESULT	7		
ID	AAG06330		
XX	AAG06330 standard; Protein; 526 AA.		
XX	AAG06330;		
DT	17-OCT-2000 (first entry)		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 3066.		
XX	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
OS	Arabidopsis thaliana.		
PX	EP1033405-A2.		
PD	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-0301439.		
XX	25-FEB-1999; 990S-0121825.		
PR	05-MAR-1999; 990S-0123180.		
PR	09-MAR-1999; 990S-0123548.		
PR	23-MAR-1999; 990S-0125788.		
PR	25-MAR-1999; 990S-0126264.		
PR	29-MAR-1999; 990S-0125785.		
PR	01-APR-1999; 990S-0127462.		
PR	06-APR-1999; 990S-0128234.		
PR	08-APR-1999; 990S-0128714.		
PR	16-APR-1999; 990S-0129845.		
PR	19-APR-1999; 990S-0130077.		
PR	21-APR-1999; 990S-0130449.		
PR	23-APR-1999; 990S-0130510.		
PR	23-APR-1999; 990S-0130891.		
PR	28-APR-1999; 990S-0131449.		
PR	30-APR-1999; 990S-0132048.		
PR	30-APR-1999; 990S-0132407.		
PR	04-MAY-1999; 990S-0132484.		
PR	05-MAY-1999; 990S-0132485.		
PR	06-MAY-1999; 990S-0132486.		
PR	06-MAY-1999; 990S-0132487.		
PR	07-MAY-1999; 990S-0132863.		
PR	11-MAY-1999; 990S-0134256.		
PR	14-MAY-1999; 990S-0134218.		
PR	14-MAY-1999; 990S-0134219.		
PR	14-MAY-1999; 990S-0134221.		
PR	14-MAY-1999; 990S-0134370.		
PR	18-MAY-1999; 990S-0134768.		
PR	19-MAY-1999; 990S-0134941.		

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141267.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

XX	17-OCT-2000	(first entry)	99US-0139750
DT	XX	Arabidopsis thaliana protein fragment SEQ ID NO: 3068.	99US-0139763
DE	XX	Protein identification; signal transduction pathway; metabolic pathway;	99US-0139817
DE	XX	hybridisation assay; genetic mapping; gene expression control; promoter;	99US-0139899
KW	XX	termination sequence.	99US-0140353
KW	XX	Arabidopsis thaliana.	99US-0140354
OS	XX	EP1033405-A2.	99US-0140695
PN	XX	06-SEP-2000.	99US-0140823
PD	XX	25-FEB-2000; 2000EP-0301439.	99US-0140991
PF	XX	25-FEB-1999;	99US-0141287
PP	XX	05-MAR-1999;	99US-0141842
PP	XX	09-MAR-1999;	99US-0142154
PP	XX	23-MAR-1999;	99US-0142055
PP	XX	25-MAR-1999;	99US-0142390
PP	XX	29-MAR-1999;	99US-0142803
PP	XX	01-APR-1999;	99US-0142920
PP	XX	06-APR-1999;	99US-0142977
PP	XX	18-APR-1999;	99US-0143542
PP	XX	16-APR-1999;	99US-0143624
PP	XX	19-APR-1999;	99US-0144005
PP	XX	21-APR-1999;	99US-0144085
PP	XX	23-APR-1999;	99US-0144086
PP	XX	28-APR-1999;	99US-0144325
PP	XX	30-APR-1999;	99US-0144331
PP	XX	04-MAY-1999;	99US-0144332
PP	XX	05-MAY-1999;	99US-0144335
PP	XX	06-MAY-1999;	99US-0144352
PP	XX	07-MAY-1999;	99US-0144632
PP	XX	11-MAY-1999;	99US-0144884
PP	XX	14-MAY-1999;	99US-0144814
PP	XX	14-MAY-1999;	99US-0145086
PP	XX	14-MAY-1999;	99US-0145088
PP	XX	14-MAY-1999;	99US-0145085
PP	XX	18-MAY-1999;	99US-0145089
PP	XX	19-MAY-1999;	99US-0145192
PP	XX	20-MAY-1999;	99US-0145218
PP	XX	20-MAY-1999;	99US-0145224
PP	XX	20-MAY-1999;	99US-0145276
PP	XX	20-MAY-1999;	99US-0145913
PP	XX	20-MAY-1999;	99US-0145918
PP	XX	20-MAY-1999;	99US-0145919
PP	XX	20-MAY-1999;	99US-0145951
PP	XX	20-MAY-1999;	99US-0146386
PP	XX	20-MAY-1999;	99US-0146388
PP	XX	20-MAY-1999;	99US-0146389
PP	XX	20-MAY-1999;	99US-0147038
PP	XX	20-MAY-1999;	99US-0147204
PP	XX	20-MAY-1999;	99US-0147302
PP	XX	20-MAY-1999;	99US-0147192
PP	XX	20-MAY-1999;	99US-0147260
PP	XX	20-MAY-1999;	99US-0147303
PP	XX	20-MAY-1999;	99US-0147416
PP	XX	20-MAY-1999;	99US-0147493
PP	XX	20-MAY-1999;	99US-0147935
PP	XX	20-MAY-1999;	99US-0148171
PP	XX	20-MAY-1999;	99US-0148319
PP	XX	20-MAY-1999;	99US-0148341
PP	XX	20-MAY-1999;	99US-0148565
PP	XX	20-MAY-1999;	99US-0148684
PP	XX	20-MAY-1999;	99US-0149368
PP	XX	20-MAY-1999;	99US-0149175
PP	XX	20-MAY-1999;	99US-0149426
PP	XX	20-MAY-1999;	99US-0149722
PP	XX	20-MAY-1999;	99US-0149723
PP	XX	20-MAY-1999;	99US-0149929
PP	XX	20-MAY-1999;	99US-0149902
PP	XX	20-MAY-1999;	99US-0149930
PP	XX	20-MAY-1999;	99US-0150566
PP	XX	20-MAY-1999;	99US-0150884


```

Db 915 ek---keivdieevkeevattlietveageeksantiteifeenleenavesnvaenl 971
QY 442 SKEKEILNELQMEATEVEKFLGRALPIKRLMALAPHNFANHETIKYVYVGSGLPGHK 501
Db 972 ekinetvntldkveetveisgennemdkaf-----fseifdnvkgigenlltg-- 1024
QY 502 RFSGWEDYEGSVIVAKICSSRRIPRYFRKSPRICGLDSRGLOLFSHG-----KHNLSPA 557
Db 1025 -----mfrsletsivigsee--kvdlnenvvssildn--ienmkedllnklenist 1072
QY 558 HSNQNVPKG--NSGCKFPKDVAMVWEKMGQFATAIVAIFILSVASKADAVDAKTCCTC 616
Db 1073 egvqetvtehveqnyvdivdpamk----dqf-----lgilneagg----- 1109
QY 617 LLKECRLELAKISNPACAAANVACLOTCNNRPDETECO-----IKCGDLFENSVDNECA 673
Db 1110 -lkemfnleedvfks-----esdvitveeikdepvqkeketvsiileemeenivdvlee-- 1163
QY 674 VSRKKCVPRKSDVGD--FPVPDPVSVLVOKFDMKDFSGKWFITRGLNPTFDADFQQLHEFH 731
Db 1164 -----ekeditdkmidaveesieiss--dskeet-----esi 1193
QY 732 TEENKLVGNLSWRIRTPDGGFFTRSAVQKVPQKYPGILYHNDNEYLYYQDDWYILSSK 791
Db 1194 kdekdvslveevqnd-----mdesvekvle-----lkmeeelmkdaveinditsk 1242
QY 792 -VENSPEYIFVYKGRNDADWDGGSVLYTRSAVLPESI--IPELOTAQAKVGRDFNTF 848
Db 1243 lieetqe-----ineveadlikmekikeka-----isedskei 1278
QY 849 IKTDNTCGPEPLVERLEKKEVERGE--RTIIEKEVEET-----EEVEKVRD-----K 893
Db 1279 idakd-----dtlekvieeedhittldevvelkdveedkievskldkeedlik 1329
QY 894 EV-----TIFSKLPEGFELORDENFURELSKEEMDVLDGLKMEATEVEKLFGRAL----- 945
Db 1330 evkeikeleseiledykelktietdileekkeiekdhfexfeeeaeieikdeadiikevs 1389
QY 946 -----PIRKMAVATHCTSPCHDIRFRFSSDDGIGRLGTRKTRINGTFLLLKIL 994
Db 1390 sleveekkleevhelkeevheisgdah-----ikg----- 1421
QY 995 PPIQSADLRTTGGSSRPLSAFRSGFSKGFIDIVPLPSKNELKELAPLLKLVGLACA 1054
Db 1422 --leeddieevddlksgildmkgdmgmd-----kesleedvttkl----- 1462
QY 1055 FLIVPSADAVDAKTCALLKGCRIELAKCIANPACAAVACLOTCN--NRPDTECOIK 1112
Db 1463 -----gerveslkdvlsalgmddeeqmktkrkaqrpkleevllk 1501
QY 1113 CGDLFENSVDNEFECASVRKKCVPRKSDLGEPFAPDPVSVLVQNFNISDPNGKWIITSG 1172
Db 1502 -----eevkeepkkittkk-----vrfdikd-----k 1524
QY 1173 NPTFADFQQLHEFHTEGDKNLKGNISWRIKTLDSGFFTRSAVQKVFQDPNPQGVLYNHD 1232
Db 1525 epkdeivevmekdieed-----veedieed----- 1551
QY 1233 NEYLYHQDDWYILSSKTEKPEDYIFVYGRNDADWDGCGAVVYRSSVLPNSIPELE 1292
Db 1552 -----ieedkvedidedidedgedekdevid-----llivqke 1583
QY 1293 KAAKSIGRDFSTFRTDNTCGPEPALVEKTEKVEGERIIVKEVER-----IEEVEE 1345
Db 1584 kriekev-----kakkkkleekveevsglkhvdevmkyvqkidevd 1626
QY 1346 KEVEKV--GRTEMTLFORLAGFNFELKODEE-----NFVRELSKEMEFLEIDKWEASE 1397
Db 1627 kevskaleskndvt-----nvllkqngdffskvknfvkkykvaapfisavaafasy 1677
QY 1398 VEKLF 1402

```

Db 1678 vvgff 1682

RESULT 12

AAR72826
ID AAR72826 standard; Protein; 2482 AA.

XX AAR72826;

XX 27-FEB-1996 (first entry)

XX Human mitotin.

XX Cell cycle; M phase; mitotin; retinoblastome; mitosis; cell growth;
inhibition.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Region 1480..1659

XX /label= internal_repeat

XX Region 1660..1839

XX /label= internal_repeat

XX WO9511309-A2.

XX 27-APR-1995.

XX 24-OCT-1994; 94WO-US12162.

XX 22-OCT-1993; 93US-0141239.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Lee W, Zhu X;

XX WPI; 1995-170229/22.

XX N-PSDB; AA086851.

XX Purified mammalian protein mitotin and agents that bind it and
inhibit its action - used to promote cell growth or to inhibit cell
division and/or proliferation

XX Claim 4; Fig 8B; 61pp; English.

XX AAR72829 is human mitotin. Mitotin is involved in the regulation of
the mammalian mitotic cell cycle. Mitotin as with E2F-1 (see AAR72824)
interacts with the retinoblastoma protein (the retinoblastoma tumour
suppressor gene product). Mitotin is first synthesised at the G1/S
boundary, it is then phosphorylated from S through M phase, and during
mitosis, is closely associated with the centromeres/kinetochores at the
mitotic spindle poles. Mitotin is necessary for a eukaryotic cell to
enter the M phase of the mitotic cell cycle and its degradation is
necessary for a cell to advance on to the next stage. Mitotin is thus
useful for controlling cell growth as overexpression of mitotin prevents
a cell from exiting the M phase.
XX An anti-mitotin antibody, antibody fragment or a phosphorylated mitotin
mutin (or nucleic acid encoding it) can also be used to inhibit cell
division which is particularly useful for the study of the cell cycle.
XX A further use is to control hyperproliferative cells, and so control
diseases such as psoriasis and breast cancer. It can also be used to
block gametogenesis of an immature gamete.

XX Sequence 2482 AA;

Query Match 2.0%; Score 147; DB 16; Length 2482;

Best Local Similarity 18.4%; Pred. No. 0.064;

Matches 240; Conservative 188; Mismatches 443; Indels 434; Gaps 60;

QY 379 LVERLEKTAEEGKLLKEAVEIEEVE-----KEVE-KVRDTEMTLQRLLEGFKELOQ 432

Db 315 lsetislekemsssiislnkreieeltqngtlknealsnqeknnliqk-sesfanyid 373

QY 433 DEENFVRELS---KEEKEIL----- 449
 Db 374 ereksiselsdqyqekililqirceetgnayedlsqykaaeknskclellnectalsce 433
 QY 450 ---NELQMEATEVEKLFGRALP--IRKLRLMALAPHSNLANHETIKYVVGSKLPQH--- 500
 Db 434 nrknele---qlkeafakehefkltafaeernqnlmleletvqqaalrsemtdngns 489
 QY 501 KRFSGWEDYFGSIWAKICSSRRIPRYFRKSPRI-----CCGLDSRGLQLFSGHGK 551
 Db 490 kseagglkqelmtikeeqnmqkqvndilqeneqlmvmktkhecqnlesepir----- 543
 QY 552 HNLSPAHSINQVPGNSGCKFPKDVAMVWKEWGWQFQAKTAIIVAIFILSVASKADAVDAL 611
 Db 544 -----nsvkresernq-cnfpqmdlevkeisldsynaqlvqleamlnrnlkelklqese 596
 QY 612 KTCCTL-----LKECRLEL-AKCISNP-----ACAAN 637
 Db 597 kekeclqhelqirgdltsnldmqsqeislglkdceidaeekeyisgphelstsqndnah 656
 QY 638 VAC-LQTCNNRPDTE--CQIKCGDLFENSVDNFNECAVSRKKCV--PRK--SDVGDF- 689
 Db 657 lqcslqttmnlkneleikicellqaekye--lvteInd---srscitatrkrmaeevgkll 711
 QY 690 ---PVPDPSVLVQKFDKDFSGKWEITRGLNP-----TFDAFDC 725
 Db 712 nevklnddsqllgelvelidpggef---geqneqhpvslapladesnsyehitlsdkev 768
 QY 726 QLHEFHEEENKLVGNLSWRINTPDGFFTRSAVQKQVQDPKYPGILYNHNDNEYLLOD-- 783
 Db 769 qmhaelqek-----flslqsehkilhdhqc--mskmselqtyvdsI 810
 QY 784 --DWYILSSKVENSPEDYIFVYKGRND-----AWDYGSGSVLYTRSAVL 826
 Db 811 kaenlvistlnrfqglvkmqlgleeglvpslssscvdpdssslsgdsffy--rall 868
 QY 827 PE---SILPELOFAAQ---KVRDRNTFTKTDNTCGPEPPL-----VERLEKKVEEGE 873
 Db 869 eqtdgmllnleqavsaancsdevfcsslqeenlrlrketpsapakgveeleslcevr 928
 QY 874 RTIJKVEEIEEEVEK---VRDKVILFSLKLFEG-----FKELQRDEENFLRELSKE 922
 Db 929 qsl-----ekleekmesgginmkneiqeqlslsseqelclrkqylseneqwqqltsv 984
 QY 923 EMDVLGDLKWEATEVEKLF-----GRALPIRKLMAVATH-----CFTS-- 960
 Db 985 tlemesklaaekkteqlslelevarlqlgglldlssrllgldtedaiggrnescdiske 1044
 QY 961 -----PCHDRIRFFSDDGIGRLGITRKRINGTFLLLKILPPI-----QSADL--- 1002
 Db 1045 htsettertphkd-vhqiokdqqdlnldiekitetgalk---ptgecsgeqsdptnye 1100
 QY 1003 -----RTTGGRS-----SRPLSAFRSFGSKGIFDVLPLPSK----- 1033
 Db 1101 ppgedktqgsseciselsfsgpnalvpmdfngqediqlrlrvketsnenlrlhviied 1160
 QY 1034 -----NELKELAPLLKLVGLVACAFILVPSADAVALKTCACLLKGCRIELAKC 1084
 Db 1161 rdrkvesllnemkeldskhlhgevlmt-----kieac-----ieleki 1199
 QY 1085 IANPACAANVACLQTCNNRPDETECOIKCGDLFENSVDNFNECAVSRKKCVPRKSDLG- 1143
 Db 1200 v-----gel-----kk-----ensldse 1212
 QY 1144 --EFPAPDPSVLVQNFISDFNGKWVITSGLNPTFDFAFCOLHEFHE-----GDNKLGVN 1197
 Db 1213 kleyfscdhqellqrvtse-----glns-----dlemhadksredidgndvakvn 1258
 QY 1198 ISWRIKTLDSGFTRSAVQKQVQDPNQPGLVYNHNDNEYLH-----YQDDWYILSSKIENK 1252
 Db 1259 dswkerfld-----venelsirirsekasiehealyleadlevvqteklclcklenenk 1310

QY 1253 P-----EDYIFVYGRNDAMDGYGGVAVVYTRSSVLPNSIIPLELEKAAKSIGRDFSTFI 1306
 Db 1311 qkvivcleelsvvtsernql---qgldtmakkttdalqglsekkmkktqelshqsecl 1367
 QY 1307 RTDNTCGPSPALVERTEKVEGERIIVKEVEIEEE---VEKEVEKYVGRTEWTLFQRLA 1363
 Db 1368 hciqvaaevkekteilqtissdvseilkkdthlqglslqlekdqslskcelengia 1427
 QY 1364 EGFNELKQDEENFVRE-----LSKEEEMFLDEIK-MEASEVEK 1400
 Db 1428 ----qinkekellvkeslqarlseidyeklnvskaleaalvek 1468
 RESULT 13
 AAW23996
 ID AAW23996 standard; Protein; 2482 AA.
 XX
 AC AAW23996;
 XX
 DT 28-MAY-1998 (first entry)
 XX
 DE Human mitosis amino acid sequence.
 XX
 KW Mitosis; phosphoprotein; mitotic cell cycle; antibody; analogue;
 KW inhibition; M phase; Antagonist; hyperproliferative cell; cancer;
 KW leukaemia; lymphoma; chromosome segregation.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Domain 258..280
 FT /note= "leucine heptad repeat"
 FT Domain 340..362
 FT Domain 564..593
 FT Domain 1387..1443
 FT Domain 1885..1962
 FT Domain 2146..2188
 FT Domain 2165..2187
 FT /note= "leucine heptad repeat"
 FT Misc-difference 2188
 FT Misc-difference 2300
 FT /label= "Bipartite targeting motif"
 FT /note= "Optionally C or G"
 FT Misc-difference 2189
 FT Misc-difference 2301
 FT Misc-difference 2303
 FT /label= "Bipartite targeting motif"
 FT /note= "Optionally A or T"
 XX
 US5710022-A.
 XX
 PD 20-JAN-1998.
 XX
 PF 24-OCT-1994; 94US-0328254.
 XX
 PR 24-OCT-1994; 94US-0328254.
 PR 22-OCT-1993; 93US-0141239.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Lee W, Zhu X;
 XX
 DR WPI; 1998-109817/10.
 DR N-PSDB; AAV09076.
 XX
 PT New isolated mitosis protein and gene - useful for, e.g. developing
 PT products for therapy and diagnosis of hyper-proliferative disorders
 PT such as cancers or psoriasis
 XX
 PS Claim 1; Column 40-52; 43pp; English.
 CC This is the amino acid sequence for mitosis, a phosphoprotein
 CC necessary for the cell to enter mitosis. The protein's degradation is

CC also necessary for the cell to advance into the next stages of mitosis.
 CC The mitosis protein, can be used to control the growth of cells. An
 CC anti-mitosis antibody, a mutant or a non-functional analogue of mitosis
 CC can inhibit the mitotic cell cycle by preventing the cells from entering
 CC the M phase, and over expression of mitosis or its functional
 CC equivalent, would inhibit the cycle by preventing cells from leaving the
 CC M phase. Antagonists to this protein can be used to control
 CC hyperproliferative cells in, (e.g. thyroid hyperplasia, grave's disease,
 CC psoriasis, benign prostatic hypertrophy, Li-Fraumeni syndrome, breast
 CC cancer, sarcomas and other neoplasms, bladder cancer, colon cancer,
 CC lung cancer and various leukemias and lymphomas). Reintroduction or
 CC supplementation of lost mitosis function by introduction of the protein
 CC or nucleic acid encoding the protein into a cell can restore defective
 CC chromosome segregation, which is a marker of progressing malignancy.
 CC Malignant proliferation of cells can then be halted. The protein
 CC can also be used for the detection and diagnosis of hyperproliferative
 CC cells.
 XX
 SQ Sequence 2482 AA;

Query Match 2.0%; Score 147; DB 19; Length 2482;
 Best Local Similarity 18.4%; Pred. No. 0.064;
 Matches 240; Conservative 188; Mismatches 443; Indels 434; Gaps 60;

QY 379 LVERLEKTAEEGKLLKEAVEIEEVEE-----KEYE-KVRDTEMTLFORLLGFKELQ 432
 DB 315 lsetlslekemsiislnkreleltqngtikeinasinqekmliqk-sesfanyid 373
 QY 433 DEENFVRELS---KEEKEL-
 DB 374 ereksiselsdqyqkelllllqrceetngayedsqykaaeknsklecilnectslce 433
 QY 450 --NELQWATEVEKLGKRALP--IRKLALAPHNSFLANHETIKYVYGSKLPGH--- 500
 DB 434 nrknele----qlkeafakehqeftklafaeernqnlmleltvqqalrsemtdnqns 489
 QY 501 KRFSWGEDYFGSIVVAKICSSRIPIRYFKSPRI-----CCGLDSRGQLQFSGHK 551
 DB 490 kseagglkqelmtlikeeqnmqkevndllqeneqlmkmvmtkhecpulesep- 543
 QY 552 HNLSPANINQVPGKSGCKFKPDVALVWYKWKQFQAKTAIIVAIFILSVASKADAVDAL 611
 DB 544 -----hsvkeresernq-cnfpqmdlevkeislsdysnagvlqleamlnrkeiklqese 596
 QY 612 KTCCTL-----LKECRLEL-AKCISNP-----ACAAN 637
 DB 597 keceqlhelqirgdietsnlqdmqsqeisglkdceidaeekeyisgphelstsqndnah 656
 QY 638 VAC-LQTCNNRPDETE--CQIKCGDLFNSVDFNFCVSRKKCV--PRK--SDVGDG- 689
 DB 657 lqcslqtmnklinelekeicellqakeye--lvtelnd---srsecitatrkmaveevgkll 711
 QY 690 ---PVDPSPVLVQKFDKDFSGKWFITRGLNP-----TFDAFDC 725
 DB 712 nevliindsgllhelvedipggef---gcqneqhpvspladesnsyehltlsdkev 768
 QY 726 QLHFHTTEENKLVGNLSWRINTPDGFFTRSAVQKVFQDPKYPGILYNHNDNEYLQD--- 783
 DB 769 qmhaefaeleq-----flsiqsehkhlhdhqcq--msskmselqtyvds 810
 QY 784 --DWYILSSKVENSPEDYIFVYKGRND-----AWDCYGGSVLYTRSAVL 826
 DB 811 kaenlvistnlnrfggdlvkmqlglleeglvpslssscvpsdsslsldgsffy--rall 868
 QY 827 PE----SLIPELQTAQAQ-----KVGREDNTFTKTDTCGPEPPL-----VERLEKKVVEEGE 873
 DB 869 eqtdgmslilsnlegavsanqscvdevfcslsigeenlrrketpsapakgveeslecevyr 928
 QY 874 RTIKVEIEEIEEVEK---VRDKVTVLFSLKLFEG-----FKELORDEENFLRELKSKE 922
 DB 929 qsl-----ekleeknesdgimkneikeileqllsserqelclrkqylseneqwgqkltsv 984

QY 923 EMDVLDGLKMEATEVEKLF-----GRALPPIRKLMVATH-----CFTS--- 960
 DB 985 tlemeskiaaekktqeqslslevarlqlqldlssrsligdtedaigqnescdiske 1044
 QY 961 -----PCHDRIRFFSSDDGIGRLGTRKIRINGTFLKILPPI-----QSADL--- 1002
 DB 1045 htsettertpkhd-vhqicdkdaqdinldiekitegalk---ptgecsqeqspdtneye 1100
 QY 1003 -----RTTGGRS-----SRPLSAFSGSGFKGFIIVPLPSK----- 1033
 DB 1101 ppgedktgsgseciselsfsgpnalvpmdfingqedlhnqlrvketsnenrllhwied 1160
 QY 1034 -----NELKELTAPLLLLKLVLAACAFILVPSADAVDALTKACALLKGCRIELAKC 1084
 DB 1161 rdrkvesillnemkeldsklhqevqlmt-----kieac-----ieleki 1199
 QY 1085 IANPACAANVACLOTCCNNRPDETECCQIKCGDLFNSVDFNFCVSRKKCVPRKSDLG- 1143
 DB 1200 v-----gel-----kk-----ensdlse 1212
 QY 1144 --ERPAPDPSVLQNFNIDFNGKWIYITSGLNPTFDAPDCOLHEFHT- ---GDNKILVGN 1197
 DB 1213 kleyfscdhqellqrvtse-----gins-----dlmhadksredigdnvakvn 1258
 QY 1198 ISWRIKTLDSGFFTRSAVQKVFQDPNPGVLYNHNDNEYLH-----YQDDWVILSSKIENK 1252
 DB 1259 dswkerfid-----venelsrirseseaiehealyleadlevvgtekclckdenek 1310
 QY 1253 P-----EDYIFVYVYGRNDWDGYYGAVVYTRSSVLPNSIPELEKAASIGRDFSTFI 1306
 DB 1311 qkvivcleelsvvtsernql---rgeldtmskktaldqisekmkktqieshqseci 1367
 QY 1307 RTDTCGPEPALVERIEKTVVEGERIIVKEVEIEE---VEKEVEKVGRTMTLFORLA 1363
 DB 1368 hciquaeevkekteellqtlssdvselikdthlqeklqslskdsqalslckcelenqia 1427
 QY 1364 EGFNELKODENFVRE-----LSKEMEFLDEIK-WEASEVEK 1400
 DB 1428 ----qlnkekellvkeseslqarisdyeklnvskaleaalevek 1468
 RESULT 14
 AAB18324
 ID AAB18324 standard; Protein; 1558 AA.
 XX
 AC AAB18324;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:182.
 XX
 KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide.
 OS Plasmodium falciparum.
 XX
 PN WO200025728-A2.
 XX
 PD 11-MAY-2000.
 XX
 PF 05-NOV-1999; 99WO-US26796.
 XX
 PR 05-NOV-1998; 98US-0107131.
 XX
 PA (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX
 PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX
 DR WPI; 2000-365347/31.
 XX

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 02:56:11 ; Search time 90.8 Seconds
(without alignments)
349.941 Million cell updates/sec

Title: US-09-075-375A-2

Perfect score: 7495

Sequence: 1 MALSLHTVLCREELNLYA.....MEASEVEKLFKALPIRKVR 1412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PTUS.COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	147.5	2.0	1786	4	US-08-973-462-8
2	147	2.0	2482	1	US-08-328-254-6
3	135.5	1.8	1312	2	US-08-592-126-148
4	135.5	1.8	1312	2	US-08-687-080-51
5	132	1.8	776	1	US-08-021-601-2
6	132	1.8	776	1	US-08-082-849B-2
7	132	1.8	776	5	PCT-US94-01624-2
8	124	1.7	1498	2	US-08-404-531B-28
9	124	1.7	1498	3	US-08-476-900A-28
10	124	1.7	1498	3	US-08-488-546A-28
11	124	1.7	1581	2	US-08-404-531B-6
12	124	1.7	1581	3	US-08-476-900A-6
13	124	1.7	1581	3	US-08-488-546A-6
14	119	1.6	905	2	US-08-574-959A-9
15	119	1.6	905	4	US-09-357-014-9
16	119	1.6	1135	2	US-08-574-959A-7
17	119	1.6	1135	4	US-09-357-014-7
18	117.5	1.6	1197	4	US-08-836-567-12
19	117.5	1.6	1230	2	US-08-968-542C-35
20	115.5	1.5	1130	2	US-08-460-309-2
21	115.5	1.5	1130	2	US-08-125-077-2
22	115.5	1.5	3111	2	US-08-460-309-4
23	115.5	1.5	3111	2	US-08-125-077-4
24	115	1.5	1435	2	US-08-568-459A-4
25	115	1.5	1435	2	US-08-487-826B-4
26	114.5	1.5	1579	3	US-08-755-587-184
27	114	1.5	523	2	US-08-482-728A-19

28	114	1.5	523	4	US-09-028-366-4	Sequence 4, Appli
29	112.5	1.5	3248	1	US-08-353-700-1	Sequence 1, Appli
30	112.5	1.5	3248	5	PCT-US95-16216-1	Sequence 1, Appli
31	111.5	1.5	1104	4	US-08-923-992A-4	Sequence 4, Appli
32	111.5	1.5	1130	6	5444158-2	Patent No. 5444158
33	111.5	1.5	2329	3	US-08-755-587-16	Sequence 16, Appli
34	111.5	1.5	3418	2	US-08-603-753D-4	Sequence 4, Appli
35	111.5	1.5	3418	4	US-09-099-753-4	Sequence 4, Appli
36	111.5	1.5	3418	4	US-08-986-106-4	Sequence 4, Appli
37	110.5	1.5	1128	4	US-08-923-992A-6	Sequence 6, Appli
38	110.5	1.5	1164	4	US-08-923-992A-2	Sequence 2, Appli
39	110	1.5	689	4	US-09-177-249-2	Sequence 2, Appli
40	110	1.5	689	4	US-09-061-769A-2	Sequence 2, Appli
41	109.5	1.5	1098	4	US-08-923-992A-8	Sequence 8, Appli
42	109	1.5	425	2	US-08-986-963-2	Sequence 2, Appli
43	109	1.5	3135	1	US-08-323-170B-2	Sequence 2, Appli
44	108.5	1.4	711	3	US-08-946-475-9	Sequence 9, Appli
45	108.5	1.4	711	4	US-09-340-479-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1

US-08-973-462-8

; Sequence 8, Application US/08973462B

; Patent No. 6191270

; GENERAL INFORMATION:

; APPLICANT: DAUBERSIES, PIERRE

; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES

; FILE REFERENCE: 0660-0125-0 PCT

; CURRENT APPLICATION NUMBER: US/08/973,462B

; EARLIER FILING DATE: 1998-02-06

; EARLIER APPLICATION NUMBER: PCT/FR96/00894

; EARLIER FILING DATE: 1996-06-12

; EARLIER APPLICATION NUMBER: FR 95/07007

; EARLIER FILING DATE: 1995-06-13

; NUMBER OF SEQ ID NOS: 29

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 8

; LENGTH: 1786

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Polypeptide

US-08-973-462-8

Query Match 2.0%; Score 147.5; DB 4; Length 1786;
Best Local Similarity 18.1%; Pred. No. 0.0031;
Matches 207; Conservative 155; Mismatches 374; Indels 409; Gaps 50;

QY 347 SIIPNLQK-AAKSVGRDFN-----NFTTNSCGPEPLVE-----ELEKTAERG 390

Db 858 TILENVETTAESVTFNSLIEIQENTITNDTI---EEKLEHENVLSALENTQSEE 914

QY 391 EKKLKEAVEIEEVEKEV-----EKVRTMTLFORLLEGKELQOD-----EENFREL 441

Db 915 EK---KEVIDVIEEVEKEVATTLIETVEQAEKSANTITEIFENLEENAVESNVAENL 971

QY 442 SKEEKELNELQMEATEVEKLFGRALPIKRLMALAPHNFLANHETIKYYVGSKLPGHK 501

Db 972 EKLNETVFTVLDKVEETVEISGESLENEMDKAF-----FSEIFDNVKGIOENLLTG-- 1024

QY 502 RFSWGWDYFGSIVVAKICSSRRIPRYFRKSPRICGLDSRGLQLFSHG-----KHLSPA 557

Db 1025 -----MFSIETSIVTQSEE--KVDLNNVSVSSILDN--IENKGLLNKLNENIST 1072

QY 558 HSIQNVKPG-NSGCKFPKDKVAMVWEKQFAKTAIVAIFILSVASKADVALKTC 616

Db 1073 EGQVETVTEHVEQNVVVDVDPAMK-----DQF-----LGLNEAGG----- 1109

Qy 617 LLKCRLELAKISNPACAAVACLOTCNNRPDETECO---IKCGDLFENSVDVEFNECA 673
Db 1110 -LREMFFNLEVDVKS---ESDVTVEIDKDPVQKEKETSIIIEEMENIVDVLEE-- 1163
Qy 674 VSRKKCVPRKSDVDG--FPVPDPFSLVQKDFMDKSGKWFITRGLNPTDFADFCOLHEFH 731
Db 1164 -----EREDLTDKMIDAVESESIESS--DSKEET-----ESI 1193
Qy 732 TEENKLVGNLSWRITPDGFFTRSAVOKFVQDPKPGIILYNDHNEVLLYQDDWYILSSK 791
Db 1194 KDKEDVSLVVEEQDND---MDESVEKYLE---LKNMEELMKDAVEINDITSK 1242
Qy 792 -VENSPEYIFVYKGRNDADWDGVSGLVTRSAVLPESI--IPELQTAQKVGROFNFT 848
Db 1243 LIETQE-----LNEVEADLIKMEKIKELEKA---LSEDSKEI 1278
Qy 849 IKTDNTCGPEPLVERLEKKEVEGE--RTIIEVEEI-----EEVEKVRD-----K 893
Db 1279 IDAKD-----DTLEKVEIEEHDTITTLDEVVELKDVDEEKIEKVSOLKDEEDILK 1329
Qy 894 EV---TLFSKLFPGFELQORDEENFLRELSEKMDVLDGLKMEATEVEKLFGRAL---- 945
Db 1330 EVKEIKESEILEYKELTIETFDILEEKEKEIKOHFEKFEAEAEIKOLEADILKEVS 1389
Qy 946 -----PIRKMAVATHCFTSPCHDRIRPFSSDDGIGRLGITRKRINGTFLKIL 994
Db 1390 SLEVEEKKLEEVHELKEEVEHIISGDAH-----IKG----- 1421
Qy 995 PPIQSADLRITGGSSRPLSAFRSGFSKGFIDIVPLPSKNEKELKELTAPLLKLVGLACA 1054
Db 1422 --LEEDDLEEVDDLKGSITLDMKGMELGDM-----KESLEDVITKL----- 1462
Qy 1055 FLIVPSADAVDALATCACLKLGCRIELAKCIANPACAAVACLOTCN--NRPDTECOIK 1112
Db 1463 -----GEVESLKDVLSSALGMDDEQMKTRKKAQRPKLEEVLLK 1501
Qy 1113 CGDLFENSVDVEFNECAVSRKKCVPRKSDILGEFPAPDPVSLVQNFNSDFNGKWIITSL 1172
Db 1502 -----BEVKEPCKKITKKK-----VRFDIKD-----K 1524
Qy 1173 NPTDFADFCOLHEFHTECDNKLGNISWRIKTLDSGFFTRSAVOKFVODPNQPCVLYNHD 1232
Db 1525 EPKDEIVEEMKDEIDED-----VEEDIEED----- 1551
Qy 1233 NEYLHQDDWYILSSKIEKNKPEDYIFVYGRNDADWDGVSGLVTRSAVOKFVODPNQPCVLYNHD 1292
Db 1552 -----IEEDKVEDIDEDIDEDIGEDKDEVID-----LIVQKE 1583
Qy 1293 KAASIGRDFSTFTINTDTCGPEPALVERIEKTVEEGERIIVKEVEE-----IEEVE 1345
Db 1584 KRIEKV-----KAKKKLEKKEVEEGVSLKKHVDVEMKVVQKIDKEVD 1626
Qy 1346 KEVEKV--GRTEMTLFORLAEFGNELKODEB-----NFVRELSEKMEFIDEITKMEASE 1397
Db 1627 KEVSALESKNDVT-----NVLKQNDQFFSKVRNFVKYKVFAPFISAVAAAFASY 1677
Qy 1398 VEKLF 1402
Db 1678 VVGFF 1682

RESULT 2

US-08-328-254-6

; Sequence 6, Application US/08328254

; Patent No. 5710022

; GENERAL INFORMATION:

; APPLICANT: Zhu, Xueliang

; APPLICANT: Lee, Wen-Hwa

; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,254
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/141,239
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 1191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2482 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-328-254-6

Query Match 2.0%; Score 147; DB 1; Length 2482;

Best Local Similarity 18.4%; Pred. No. 0.0058;

Matches 240; Conservative 188; Mismatches 443; Indels 434; Gaps 60;

Qy 379 LVRELEKTAEEGKLLIIEAVEIEEVE-----KEVE-KVRDTEMTLQRLLEGKELQO 432
Db 315 LSETLSLEKKEMSSISLNKREIELTQENGTLKETAASLNQKRMNLIQK-SSEFANYID 373
Qy 433 DEENFVRELS---KEEKEIL----- 449
Db 374 EREKSISELSDOYKQKELLILLOQCEGTNAYEDLSQYKAAQKNSKLECLLNECTSLCE 433
Qy 450 ---NELQMPATEVEKLFGRALP--TRKRLMALAPHSNLANHETIKYVYVSKLPCH---- 500
Db 434 NRKNELE---OLKEAFAKEHOFELTKLAFABERNQNLMELETYVQOALRSEMFTDNQNS 489
Qy 501 KRFSGWEDYFSGIIVVAKTCSRRIPRYFRKSPRI-----CCGLDSRGLQFSGHK 551
Db 490 KSEAGGLKOEIMTLKEEQNKMKQKQEVNDLLOENFQMLKVMKTKHECONLESEPIR----- 543
Qy 552 HNLSPAHSINQNPVKNQSGCKFPKDVAMLVWEKVGQFATAIVAFILSVASKADAVDAL 611
Db 544 -----NSVKERESERNQ--CNFKPQMDLEVKESILDSVNAQLVQLEAMLRNKLKQSE 596
Qy 612 KTCCTCL-----LKBCRLLEL-AKCTSNP-----ACAAN 637
Db 597 KEKECLOHELOQTIRGDLTSNLDQMSQBSISGLKCEIDAEKYSIGPHELSQNDNAH 656
Qy 638 VAC-LQTCNNRPDETE--CQIKCGDILFENSVDVEFNECAVSRKKCV--PRK--SDVGDFF- 689
Db 657 LQCSLQTTNKNLELEKICEILQAEKYE--LVTELND---SRSECITATRKMAEEVSGKIL 711
Qy 690 ----PVPDFSVLVQKDFMDKDFSGKWFITRGLNP-----TFDAFDC 725
Db 712 NEVKILNDSDGLHGLVEDIPGGEF---GEQPNQHPVSLAPLDESNSYEHLLTSDKEV 768
Qy 726 QLHEFTEENKLVGNLSWRITPDGFFTRSAVOKFVQDPKPGIILYNDHNEVLLYQD-- 783
Db 769 OMHFAELQEK-----FLSLQSEHKILHDOHCQ--MSSKMSSELQTYVDSL 810
Qy 784 --DWYILSSKVENSPEDYIFVYKGRND-----AWDGYGGSVLYTRSAVL 826

Db 811 KAENLVSLNLRNFGDLVEMQGLGSLVPSLSSSCVPDPSSLSLSSGSSFF--RALL 868
*QY 827 PE-----SIIPELOTRAAQ-----KVRGRDNFTTKTONTGCPPEPL-----VERLEKKVEEGE 873
Db 869 EOTGMSLSLNLGAVSANOCSDVEFCSSLOEENLTKETPSAPAKGVEELESCEVYR 928
QY 874 RTIIEVEIEEVEEK-----VRDKVTLFSKLFEG-----FKELORDEENFLRELSKE 922
Db 929 QSL-----EKLEKESQIMANKELIQELQLSSERELDCLRKQYLSSENFQWQOKLTSV 984
QY 923 EMDVLGLMKATEVEKLF-----GRALPIRKLMAVATH-----CFTS-- 960
Db 985 TLEMESKLAEEKQTEQLSLEVARLQGLDLSSRLSLGIDTFDAIOGNESCDISKE 1044
QY 961 -----PCHDRIRFSSDDGIGRIGTRKINGRFTLLKILPPI-----OSADL--- 1002
Db 1045 HTSETTEPKHD--VHOICDKDAQDNLNDIKITGTALK--PTGECSEGEQSPDTNYE 1100
QY 1003 -----RTTGRS-----SRPLSAFRSGFSKGIFDIVPLPSK----- 1033
Db 1101 PPGEDKTQGSSECSISELFSFGPNALVPMDFLGNQEDTHNLQLRVKETSNENLRLHHVIED 1160
QY 1034 -----NELKELTAPLLKLVGLVACAFLLIIPVSADAVDAKTCACLLKGRIELAKC 1084
Db 1161 RDRKVESLLNEMKELDKLHLEQVQMT-----KIEAC-----IELEKI 1199
QY 1085 IANPACANVACLOTCNNRPDETECOJLKGDLFENSVDENECVSRKKCVPRKSDLG- 1143
Db 1200 V-----GEL-----KK---ENSDLSE 1212
QY 1144 -----EPPAPDPSPVLQVONISDFNGKWIYITSGLNPTFDADFQCLHEFHE-----GDNKLGVN 1197
Db 1213 KLEFSCDQELQORVETSE-----GLNS-----DLENHADSSREDIGDNVAKVN 1258
QY 1198 ISWRIKTLGFFTRSAVKFQVQDPNPGVLYNHNDNEYLH-----YQDDWYIILSSKIENK 1252
Db 1259 DSWKREFLD-----VENELSRIRSEKASIEHEALYLEADLEVQVTEKLCLEKDNENK 1310
QY 1253 P-----EDXIFVYVYGRNDWDGCGAVVYTRSSVLPNSIIPELEKAAKSIGRDFSTFI 1306
Db 1311 QKVIVCLLEELSVVTSERNQJ-----RGELDTMSKKTALDQLSEKMKETQOELSHOSECL 1367
QY 1307 RTDTCGPPEPALVERIKTVEEGERIIVKEVEEIEE-----VEKEVEKVGRTMTLFORLA 1363
Db 1368 HICQVAEAEVKEKTELLQTLSSDVSELLKDKTHLQEKLOSLQSLQSLSTKCELENQIA 1427
QY 1364 EGFNELQODEENFVRE-----LSKEEMEFLEDEIK--MEASEVEK 1400
Db 1428 ---OLNKEKELLVKESESLOARLSSESDYEKLVNSKALEAALVEK 1468

RESULT 3

US-08-592-126-148
; Sequence 148, Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
; US-08-592-126-148

Query Match 1.8%; Score 135.5; DB 2; Length 1312;

Best Local Similarity 19.7%; Pred. No. 0.02;

Matches 124; Conservative 95; Mismatches 218; Indels 191; Gaps 24;

QY 382 RLEKTAEEGE---KLLIKEAVEIEEVEKEVEKVRDTEMTLQRLLEGFKELQODEENF 437
Db 398 KLVREQEAGEAKTQNMNDFAEKETLKQKQIDEIRD-KKTGLGRIELKSEI----- 449
QY 438 VRELSKEEKEIINELQMEATEVEKLFGRALPRTKRLMALAPHSNFI-----ANHETIKY 491
Db 450 ---LSKKQ---NELKNVYELQQLGEGSSDRILELQELIKAEARELSKAENKSNVETLKM 502
QY 492 YVGSKLPGHKRFSGWEDYFGSIVVAKICSSRRIPRYPRKSPRICCGDLSRGLQFSHGK 551
Db 503 EVIS-LONEK-----ADLORTLK-----LDQEMEQL-NH-- 530
QY 552 HNLSPAHSINQVPGKNGSCCKPKDVALMVWBEKWQGFATAIATVIFILSVASKADAVDAL 611
Db 531 -----HTTTRQMEMLTKDKADKDEQIRKIKRSUSDLSLLGYF---PNKKQLEDWL 580
QY 612 KTCCLLRECELELAKCISNPACANVACLOTCNNRPDETECOIK-----CGDLFE 662
Db 581 HSKSKEINOTRDLAKL--NKELASSEQNKHNINNELKRRERQSLSYEDKLFVCGSQDF 638
QY 663 NSWVDFEFCVSRKKCVPRKSDVGFVPDPVSLVQKFMKDFSGKWFITRGLNPTFDA 722
Db 639 ESDLDRLKE-----EIEKSKQAMLAGATAVYSQ 668
QY 723 FDCQLHEFTEENKLVGNLSWRIRTPDGGFTTRSAVKFQVQDPKYPGILYNHNDNEYLQ 782
Db 669 FITQL---TDENQCCPVCQRV---FQTEAELQEVISD----- 700
QY 783 DDWYILSSKVENSPEDYIFV--YKGRNDAMDGYCGSVLYTRSAV-LPESIIPELOTRAAQ 839
Db 701 -----LQSKRLAPDKLSTSELSAKKRRDEMGLVPMQOSIIDLKEKEIPELRNKLQ 755
QY 840 KYGRDFTFIKTD-----NTCCPEPPLVERLEKKVEEGERTIIEK-- 879
Db 756 NVNRDIQR-LKNDIEEQETLLGTIMPEESAKVCLTDVITMERQOMELKDVYERKTAQAA 814
QY 880 -----VEEIEEVEKVRDKVTLFSKLFEGFKELQODEENF-----LRELSKEE 923
Db 815 KLQGLDRTVOQVNOEKQKHLDITVSSKIELNRKLIQDQEQEQIOHKLKSTTNELKSEK 874
QY 924 MDVLDGLKME-----ATEVEKLF 941
Db 875 LOISTNLQRRQOLERQTVELSTEVSQSLY 902

RESULT 4

```

US-08-687-080-51
; Sequence 51, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TRANS. OF RAD50 cDNA (SEQ. 54), NT.
; INDIVIDUAL ISOLATE: 389 TO 4324
US-08-687-080-51

Query Match 1.8%; Score 135.5; DB 2; Length 1312;
Best Local Similarity 19.7%; Pred. No. 0.02;
Matches 124; Conservative 95; Mismatches 218; Indels 191; Gaps 24;

QY 382 RLEKTAEEGE---KLIKKAVEIEEVEEVEKVRDTEMTLFLRLLGPKELQDDENF 437
Db 398 KLVREGEQGEAKTNLMNDFAEKTLKQIDRID-KKTGLRIELKSEI----- 449
QY 438 VRELKSEKEITLQELQATEVEKLFGRALPIKRLMALAPHNLF-----ANHETIKY 491
Db 450 ---LSKKQ---NELKNVYKLEQLEGSDRIELELQDELKABRELSKAENSNVETLKM 502
QY 492 YVSKLPCHKRFSGWEDYFGSIIVVAKICSSRRIPRYFRKSPRICCGLSRGLQFLSHGK 551
Db 503 EVIS-LQNEK-----ADLDRFLRK-----LDQEMEQ-L-NH-- 530
QY 552 HNLSPAHISINQVPGNSGCKFPKPDVAMVWEKMGQFAKTAIVAFILSVASKADAVDAL 611
Db 531 -----HTTTRTQEMLTQDKADKDEQIRKIKSRHSDELTSLLGYF-----PNKQLLEDWL 580
QY 612 KTCCLLKCECLRELAKISNPACAAVACIQTCCNRRPDETECQIK-----CGDLFE 662
Db 581 HSKSKEINQTRDLRAKL--NKELASSEQNKNHINNELKRKEQLSSYEDKLFVDCGSQDF 638
QY 663 NSVVDENECASVRKCKVPRKSDVGFPPDPSVLQKFDKDFSGKWFITRGLNPTFDA 722

; Sequence 51, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TRANS. OF RAD50 cDNA (SEQ. 54), NT.
; INDIVIDUAL ISOLATE: 389 TO 4324
US-08-687-080-51

Query Match 1.8%; Score 132; DB 1; Length 776;
Best Local Similarity 17.2%; Pred. No. 0.017;
Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

Db 639 ESDLRLKE-----EIEKSSKORAMLAGATAVYSQ 668
QY 723 FDCQLHEFHTEENKLVNLSWRITPPGGFTTSRAVOKFVQDPKYPGILYNHIDNEYLLYQ 782
Db 669 FITQL---TDENQSCCPVQQRV-----FQTEAEQLQEVISD----- 700
QY 783 DDWYILSSKVENSPEDYIFV--YKGRNDAMDGYGSSVLYTRSAV-LPESITPELOTAQAQ 839
Db 701 -----LQSKRLAPDKLSTESLKKKREDEMLGLVPMRQSIIDLKKEIPELRNKLQ 755
QY 840 KVRGRDNFTFIKT-----NTGPPPEPLVERLEKKVVEGERTIIE-- 879
Db 756 NVNRDIQR-LKNDIEEQETLLGTIMPEESAKVCLTDTVIMERFQMKDKVERKIAQAAA 814
QY 880 -----VEETEEVEKVRDKETVLFSLKFGFKELQDRDENF-----LRELSKEE 923
Db 815 KLGQIDLDRTVQQVNOEKOQKHLDTVSSKIELNRKLIQDQEQIOHLSKTTNELASEK 874
QY 924 MDVLDGLKME-----ATEVEKLF 941
Db 875 LQISTNLQRROOLEEQTVELSTEVSQSLY 902

RESULT 5
US-08-021-601-2
; Sequence 2, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppala, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-021-601-2

```


Db 511 AGYLENG-----KLILORNGLEIKVO-----533
QY 1223 NOPGVLYNHDEYLYHODWYILSSKIENKPEDYIEFYVYGRNDAMDGYGGAVVYTR--S 1280
Db 534 ----IIQSEKEYIRI-DAKVVPSKIDTKIOB-----AQLNINOENKALGLPKYTKLIT 584
QY 1281 SVLPNSIIPLEKAASIGRDFSTFTDNTGCPPEPALVERIEKTVERGE-RIIVKEV-- 1337
Db 585 FNVHNRYSNIVESAYLILNEWKNIIQSD-----LIKVTNYLVDCNGRPFVFTDITL 636
QY 1338 -----EEIEEEVEKEVEKVGRTMTLFORLAEGFNEKODEENFVRE-----1379
Db 637 PNTAEQYTHODEIYEQVHSGLYVPESRILLHGPSKV-ELRNDSEGFIERHGHAVDDY 695
QY 1380 ----LSKEME-----FLDEIKMEASEV 1398
Db 696 AGYLLDKNQSDLVNTNSKKFIDIFKEEGSNL 725
RESULT 7
PCT-US94-01624-2
Sequence 2, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Slough, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND KEOURIE and CREW
STREET: Stewart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 776 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-2

* Query Match 1.8%; Score 132; DB 5; Length 776;
Best Local Similarity 17.2%; Pred. No. 0.017;
Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

QY 733 ENKLVGNLSWRITPDGFTTRSAVQKFQDPKYPGILYNHD---NEYLLYQDDWY--I 787
Db 72 EMYKAIGG---KIYIVDGDITPKHSLEALSBDKKIKIDYKDKALLHEHYVYAKEGEV 128

QY 788 LSSKVENSPEDYIEFYVYKGRNDAMDGYGGSVLYTRSAVLPEESIIPELOTAQAQKVGDRDNT 847
Db 129 L---VIQSEDIYVENTEALN-----VYIEIGKILSRDILSKINQPYQKFLDLVNT 176
QY 848 FIKDNTGCG-----PEPPLVERLEKKVEEGERITIK-----878
Db 177 IKNASDSQDQLLFTNQLKEHPTDFSVEFLBONSNEVQEVFAKAFAYYIEFPQHRDVLQY 236
QY 879 -----EVEIEEEVEKVRDKVTLFSLKLFEGFKE-----907
Db 237 APEAFNYMDKNEQENINLSLEELKQDRLMSRYEKWEKIKQHYQHSWSLSSEEGCILKKL 296
QY 908 ---LQDDEENFLRELSKEEMDVLDGLKMEATEV---EKLFGRALPI-----RKL 950
Db 297 QIPIEPKDDIIHLSLQSEKELLAKRIQIDSDSDFLSTEKEFLKLLQIDIRDSISEEKEKEL 356
QY 951 MAVATHCTFSCHDRIRFFSSDDGIGRGLGITRKRRLNGITFLLLKILPPIQSAD---LRTTG 1006
Db 357 L-----NRIQVDSSNP-----LSEK--EKEFLKLLKLDIQPYDINORLQDGTG 396
QY 1007 GRSSRPL-----SAFRSGFGSKGIFDIPVLPSPKNEKELTAPILLLKLVG 1049
Db 397 GLIDSPSINLDVRKQYKRDIONIDALLHQSIGSTLYNKIYLYENMINNLATL-----450
QY 1050 VLACAFILVPSADAVDAKLTCACLLKGCRIELAKCIANPACAAVACLOTNNRPDETEC 1109
Db 451 -----GADLVDS-----458
QY 1110 QIKCGDLFENSVDDE--FNECAVSRKKKCVPRK---SDLGEPAPDPSPVLVQNFNISDFNG 1164
Db 459 -----DNTKINRGIFNEFKFKFKYSISSNYMIVDNERPALDNEKWKRIQLSPDTR 510
QY 1165 KWTYITSGLNPTFDAPDCQLHEFHTEGDNKLV--GNISWRIKTLDGSGFTFRSAVOKFVODP 1222
Db 511 AGYLENG-----KLILQRNIGLEIKDVQ-----533
QY 1223 NOPGVLYNHDEYLYHODWYILSSKIENKPEDYIEFYVYGRNDAMDGYGGAVVYTR--S 1280
Db 534 ----IIQSEKEYIRI-DAKVVPSKIDTKIOB-----AQLNINOENKALGLPKYTKLIT 584
QY 1281 SVLPNSIIPLEKAASIGRDFSTFTDNTGCPPEPALVERIEKTVERGE-RIIVKEV-- 1337
Db 585 FNVHNRYSNIVESAYLILNEWKNIIQSD-----LIKVTNYLVDCNGRPFVFTDITL 636
QY 1338 -----EEIEEEVEKEVEKVGRTMTLFORLAEGFNEKODEENFVRE-----1379
Db 637 PNTAEQYTHODEIYEQVHSGLYVPESRILLHGPSKV-ELRNDSEGFIERHGHAVDDY 695
QY 1380 ----LSKEME-----FLDEIKMEASEV 1398
Db 696 AGYLLDKNQSDLVNTNSKKFIDIFKEEGSNL 725

RESULT 8

US-08-404-531B-28
Sequence 28, Application US/08404531B
Patent No. 5863724
GENERAL INFORMATION:
APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonyleurea Receptor
Patent No. 5863724
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
ADDRESS: No. 5863724ris
STREET: One Liberty Place 46th. Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

Db 846 STPMIVLDDPPSALDVHLSHDLMQAGILELLRDDRRTVVLVTHKLOYLPHA-DWIIAMK 904
 QY 1248 KIENKPEDYIFYVYGRNDAMDYGAGVYTRSSVLPNSIIPELEKAAKSIGRDFSTFIR 1307
 Db 905 DGTIQREGTLKDFORSECQLFEHW-----KTLNMQDQOELEK-----ETVME 946
 QY 1308 TDNTCGPEPALVERIEKTVVEGERIIVKEVEEIEEVEKE 1347
 Db 947 RK---APEPS--QGLPRAMSSRDGLLDEDEEEEAASE 981
 RESULT 10
 US-08-488-546A-28
 ; Sequence 28, Application US/08488546A
 ; Patent No. 6054313
 ; GENERAL INFORMATION:
 ; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
 ; APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
 ; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
 ; Patent No. 6054313
 ; NUMBER OF SEQUENCES: 49
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
 ; ADDRESSEE: No. 6054313ris
 ; STREET: One Liberty Place 46th. Floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/488,546A
 ; FILING DATE: 07-JUNE-1995
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/404,531
 ; FILING DATE: 15-MARCH-1995
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Beardell, Lori Y.
 ; REGISTRATION NUMBER: 34,293
 ; REFERENCE/DOCKET NUMBER: BYLR-0026
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-568-3100
 ; TELEFAX: 215-568-3439
 ; INFORMATION FOR SEQ ID NO: 28:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1498 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-488-546A-28

Query Match 1.7%; Score 124; DB 3; Length 1498;
 Best Local Similarity 18.8%; Pred. No. 0.24;
 Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;
 QY 868 KVEGERTIIEVEIEEVEKVRDKEVFLSKLFEFGK--ELQDDEENFLRELSKEEMD 925
 Db 480 KLSQAQRTTL-----EYSNEELKQT---NEMLRGIKLLKYAWENIFCSRVKTRRK 528
 QY 926 VLDGLKMEFA--TEVEKLGFRALPRKLMVAVATHCFTSPCHDRIRFFSSDD-----GIGRL 978
 Db 529 EMTSLRAFVNTSIFMNTAIPAAVLITFV-----GHVSFFKESDFSVAFAFL 580
 QY 979 GITRKRINGTFL-----KILPFIQ-----SADLRITGGRSRRLSAFRGFSKG 1023
 Db 581 SLFLHVLTPFLLSVVRSTVKALYSVKLSSEFLSSAIREEOCAPAPQQA-----G 636

QY 1024 IFDIVPL-----PSKNELKELTAPILLKLVGLVCAFLVPSADAVDAKTCACLLKG 1076
 Db 637 KYQAVPLKVVNRKRPAAREEVRDLGLPLQ-----RLTFSTDG-DADNFCVQLIIG 684
 QY 1077 C-----RIELAKCIANPACAAANVACIQTCNNRPDETEFCQIKCGDLF 1117
 Db 685 FFTWPDGIPTLNITIRIPRQGLTMIVGVGCGKSSILLATLG-----EMQKVSQAVF 738
 QY 1118 ENSVVDENECASVRKKCVPRKSLGFEPPADP-----SVLVQFNFI-SDF 1162
 Db 739 WNSLPD-----SEGRRPQQPRAGDSGRFCQEQPCGYASQKWPMLLNATVEENITFESPF 793
 QY 1163 NGKWY-----ITSGLNPTFDADFCQLHEFHTEGDNKLVGNISWRIKT----- 1204
 Db 794 NKQRYKWKVIEACSLQPDIDIL-----PHGDTQIGERCINLSTGGQRPDQCRPEPST 845
 QY 1205 -----LDSGF-----FTRSAVQKFEVQDPNPGVLYNHEDNEVLYHQDDWYILSS 1247
 Db 846 STPMIVLDDPPSALDVHLSHDLMQAGILELLRDDRRTVVLVTHKLOYLPHA-DWIIAMK 904
 QY 1248 KIENKPEDYIFYVYGRNDAMDYGAGVYTRSSVLPNSIIPELEKAAKSIGRDFSTFIR 1307
 Db 905 DGTIQREGTLKDFORSECQLFEHW-----KTLNMQDQOELEK-----ETVME 946
 QY 1308 TDNTCGPEPALVERIEKTVVEGERIIVKEVEEIEEVEKE 1347
 Db 947 RK---APEPS--QGLPRAMSSRDGLLDEDEEEEAASE 981

RESULT 11
 US-08-404-531B-6
 ; Sequence 6, Application US/08404531B
 ; Patent No. 5863724
 ; GENERAL INFORMATION:
 ; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
 ; APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
 ; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
 ; Patent No. 5863724
 ; NUMBER OF SEQUENCES: 49
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
 ; ADDRESSEE: No. 5863724ris
 ; STREET: One Liberty Place 46th. Floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/404,531B
 ; FILING DATE: 15-MAR-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Beardell, Lori Y.
 ; REGISTRATION NUMBER: 34,293
 ; REFERENCE/DOCKET NUMBER: BYLR-0003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-568-3100
 ; TELEFAX: 215-568-3439
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1581 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-404-531B-6

Query Match 1.7%; Score 124; DB 2; Length 1581;
Best Local Similarity 18.8%; Pred. No. 0.26; Indels 178; Gaps 26;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

QY 868 KVEEGERTIIKEVEIEEVEKVRDEKVTFLSKLFEFGK--ELQORDEENFLRELSKEEMD 925
Db 480 KLSQAQRTTL-----EYSNERLKQT---NEMLRGKLLKLYAWENIFCSRVKTRRK 528
QY 926 VLDGLKMEA--TEVEKLFGRALPIRKLMAVATHCFTSPCHDRIRFFSSD-----GIGRL 978
Db 529 EMTSLRAFAVYTSISIFMTAIPAAVLITFV-----GHVSFFKESDFSVAFAFL 580
QY 979 GITRKRINGTFL-----KILPPIQ-----SADLRTTGGSSRPLSAFRSGFSKG 1023
Db 581 SLFHILVTPFLLSVVRSTVKALVSOKLSEFLSSAEIREEOCAPREPAQQA-----G 636
QY 1024 IFDIVPL-----PSKNELKELTAPLLKLVGLACAFIVPSADAVDAKLTACCLLK 1076
Db 637 KYQAVPLKVVNRKRPAREEVRDLGLPLQ-----RLTPSTDG-DADNFCVQIIG 684
QY 1077 C-----RIELAKCIANPACAAVACLOTNNRPDETECQIKGDLF 1117
Db 685 FTTWTPDGIPLTNTIRIPRGQLTMIQVGVCGCKSSLLATLG-----EMOKVSGAVF 738
QY 1118 ENSVVDDEFNECAVRKCKVPRKSDLGEPAPDP-----SVLVQNFNI-SDF 1162
Db 739 WNSLPD-----SEGRRPQOPRAGDSGRFCQEQPCGYASOKPWLNLATVEENITFESPF 793
QY 1163 NGKWY-----ITSGLNPTDAFDCQLHEFTGDKNLVGNISWRIKT-----1204
Db 794 NKORYKMWIEACSLQPDIDIL-----PHGDTQIGERGINLSTGGQRPDQCRPEPST 845
QY 1205 -----LDSGF-----FTRSAVOKFVODPNQPGVLYNHDNEXLHYQDDWYLLSS 1247
Db 846 STPMIVFLDDPFSALDVHLSHLMQAGILELRDDKRTVVLVTHKLYLPHA-DWIIAMK 904
QY 1248 KIENKPEYIFVYGRNDAMDYGAVVYTRSSVLPNSIIEPELEKAASIGRDFSTFIR 1307
Db 905 DGTIQREGTLKDFORSECOLFEHW-----KTLNRQDOLEK-----ETVME 946
QY 1308 TDNTCGPEPALVERIEKTVVEGERIIVKEVEIEEVEKE 1347
Db 947 RK---APEPS---QGLPRAMSSRDGLLLDEDEEEEAASE 981

RESULT 13
US-08-476-900A-6
; Sequence 6, Application US/08476900A
; Patent No. 6031150
; GENERAL INFORMATION:
; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson
; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 6031150
; TITLE OF INVENTION: and Method of Detecting Persistent Hyperinsulinemic Hypoglycemia
; TITLE OF INVENTION: Infancy
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESS: No. 6031150 is
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,900A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800

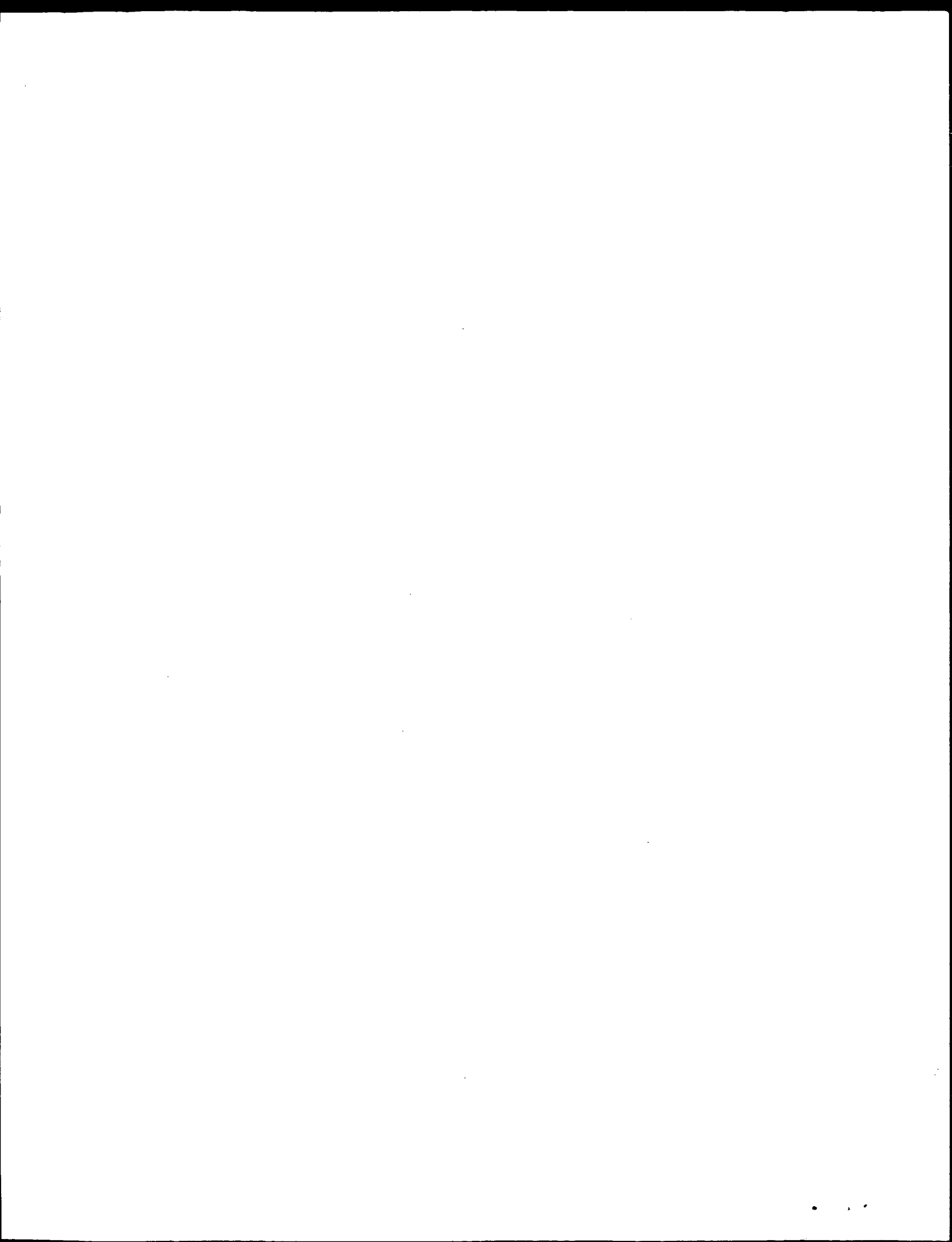
ATTORNEY/AGENT INFORMATION:
NAME: Beardell, Lori Y.
REGISTRATION NUMBER: 34,293
REFERENCE/DOCKET NUMBER: BYLR-0027
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1581 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-476-900A-6

Query Match 1.7%; Score 124; DB 3; Length 1581;
Best Local Similarity 18.8%; Pred. No. 0.26;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

QY 868 KVEEGERTIIKEVEIEEVEKVRDEKVTFLSKLFEFGK--ELQORDEENFLRELSKEEMD 925
Db 480 KLSQAQRTTL-----EYSNERLKQT---NEMLRGKLLKLYAWENIFCSRVKTRRK 528
QY 926 VLDGLKMEA--TEVEKLFGRALPIRKLMAVATHCFTSPCHDRIRFFSSD-----GIGRL 978
Db 529 EMTSLRAFAVYTSISIFMTAIPAAVLITFV-----GHVSFFKESDFSVAFAFL 580
QY 979 GITRKRINGTFL-----KILPPIQ-----SADLRTTGGSSRPLSAFRSGFSKG 1023
Db 581 SLFHILVTPFLLSVVRSTVKALVSOKLSEFLSSAEIREEOCAPREPAQQA-----G 636
QY 1024 IFDIVPL-----PSKNELKELTAPLLKLVGLACAFIVPSADAVDAKLTACCLLK 1076
Db 637 KYQAVPLKVVNRKRPAREEVRDLGLPLQ-----RLTPSTDG-DADNFCVQIIG 684
QY 1077 C-----RIELAKCIANPACAAVACLOTNNRPDETECQIKGDLF 1117
Db 685 FTTWTPDGIPLTNTIRIPRGQLTMIQVGVCGCKSSLLATLG-----EMOKVSGAVF 738
QY 1118 ENSVVDDEFNECAVRKCKVPRKSDLGEPAPDP-----SVLVQNFNI-SDF 1162
Db 739 WNSLPD-----SEGRRPQOPRAGDSGRFCQEQPCGYASOKPWLNLATVEENITFESPF 793
QY 1163 NGKWY-----ITSGLNPTDAFDCQLHEFTGDKNLVGNISWRIKT-----1204
Db 794 NKORYKMWIEACSLQPDIDIL-----PHGDTQIGERGINLSTGGQRPDQCRPEPST 845
QY 1205 -----LDSGF-----FTRSAVOKFVODPNQPGVLYNHDNEXLHYQDDWYLLSS 1247
Db 846 STPMIVFLDDPFSALDVHLSHLMQAGILELRDDKRTVVLVTHKLYLPHA-DWIIAMK 904
QY 1248 KIENKPEYIFVYGRNDAMDYGAVVYTRSSVLPNSIIEPELEKAASIGRDFSTFIR 1307
Db 905 DGTIQREGTLKDFORSECOLFEHW-----KTLNRQDOLEK-----ETVME 946
QY 1308 TDNTCGPEPALVERIEKTVVEGERIIVKEVEIEEVEKE 1347
Db 947 RK---APEPS---QGLPRAMSSRDGLLLDEDEEEEAASE 981

RESULT 13
US-08-488-546A-6
; Sequence 6, Application US/08488546A
; Patent No. 6054313
; GENERAL INFORMATION:
; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
; APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 6054313
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &

Search completed: November 6, 2001, 05:01:12
Job time: 7501 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 03:10:56 ; Search time 82.15 Seconds
(without alignments)
1309.294 Million cell updates/sec

Title: US-09-075-375A-2
Perfect score: 7495
Sequence: 1 MALSLHTVLCKEEALNYA.....MEASEVEKLFKALPIRKVR 1412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2557	34.1	478	2 T03750	violaxanthin de-ep
2	2430	32.4	462	2 T00708	violaxanthin de-ep
3	210	2.8	522	2 A84606	hypothetical prote
4	194.5	2.6	2289	2 T28677	rhodopy protein -
5	167.5	2.2	2401	2 T28676	rhodopy protein -
6	158.5	2.1	1957	2 T30877	hypothetical coile
7	156	2.1	1365	2 T30822	lmp1 protein - Myc
8	155.5	2.1	886	2 H69378	conserved hypothet
9	155.5	2.1	2829	2 A42771	reticulocyte-bind
10	155	2.1	2469	2 H36812	hypothetical prote
11	151	2.0	1939	2 T18372	repeat organellar
12	148.5	2.0	1935	2 A59286	myosin heavy chain
13	146	1.9	1935	1 A37102	myosin beta heavy
14	146	1.9	2748	2 S57976	nuclear migration
15	144.5	1.9	839	2 S54174	DNA topoisomerase
16	144	1.9	2166	2 G70163	hypothetical prote
17	143.5	1.9	1558	2 B71603	RESA-H3 antigen pF
18	142.5	1.9	880	2 F75103	conserved hypothet
19	142.5	1.9	1199	2 T29145	hypothetical prote
20	142.5	1.9	1819	2 A71928	cag island protein
21	142.5	1.9	1935	1 S06006	myosin beta heavy
22	142.5	1.9	5105	2 T32650	hypothetical prote
23	141	1.9	1002	2 C70319	nitrite reductase
24	140.5	1.9	1302	1 JC6009	surface-located me
25	140	1.9	800	2 F64508	hypothetical prote
26	139.5	1.9	1025	2 S54044	probable membrane
27	139	1.9	2712	2 T05113	hypothetical prote
28	137.5	1.8	1875	2 S38173	myosin-like protei
29	137	1.8	1109	2 A40801	phosphoprotein pho

30 137 1.8 1676 2 E71410 probable centromer
31 136.5 1.8 3724 2 T18427 hypothetical prote
32 136 1.8 1127 2 T28317 ORF MSV156 hypothe
33 135 1.8 3660 1 S02041 dystrophin, muscle
34 134.5 1.8 1937 2 I38055 myosin heavy chain
35 134 1.8 978 2 A70387 conserved hypothet
36 134 1.8 1639 2 S05603 major merozoite su
37 134 1.8 2261 2 T20978 hypothetical prote
38 133.5 1.8 1934 2 I48153 myosin heavy chain
39 133 1.8 1624 2 T25592 hypothetical prote
40 133 1.8 1979 2 C71622 hypothetical prote
41 132.5 1.8 1156 2 B70356 chromosome assembl
42 132 1.8 809 1 JQ0032 anthrax toxin leth
43 131.5 1.8 1170 2 A72287 hypothetical prote
44 131 1.7 1156 2 E69444 chromosome segrega
45 130.5 1.7 1631 1 SAZQK1 major merozoite su

ALIGNMENTS

RESULT 1

T03750

violaxanthin de-epoxidase precursor - common tobacco

C:Species: Nicotiana tabacum (common tobacco)

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000

C:Accession: T03750

R:Bugos, R.C.; Hieber, A.D.; Yamamoto, H.Y.

J. Biol. Chem. 273, 15321-15324, 1998

A:Title: Xanthophyll cycle enzymes are members of the lipocalin family, the first ide

A:Reference number: Z15054; MUID:98288256

A:Accession: T03750

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-478 <BUG>

A:Cross-references: EMBL:U034817; NID:gl463122; PIDN:AAC50031.1; PID:gl463123

A:Experimental source: strain Xanthi; tissue-type leaf

C:Genetics:

A:Gene: TVDEL

C:Function:

A:Description: violaxanthin de-epoxidase and zeaxanthin epoxidase catalyze the additi
d in protecting the photosynthetic apparatus from excessive light
A:Note: established as member of the lipocalin family
F:1-134/Domain: transit peptide (plastid) #status predicted <TNP>
F:135-478/Product: violaxanthin de-epoxidase #status predicted <MAT>

Query Match 34.1% Score 2557; DB 2; Length 478;
Best Local Similarity 100.0%; Pred. No. 4.4e-142;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 MALAPHSNLANHETIKYVVGSKLPGHKRFSWGWDYFGSIYVAKICSSRRIPRYFRKSP 533

|||||

Db 1 MALAPHSNLANHETIKYVVGSKLPGHKRFSWGWDYFGSIYVAKICSSRRIPRYFRKSP 60

|||||

QY 534 RICCGLDSRGLQFSGHKHNLSPAHSINQNVKPGNSGCKFPKDVAMVWEKQFAKTAI 593

|||||

Db 61 RICCGLDSRGLQFSGHKHNLSPAHSINQNVKPGNSGCKFPKDVAMVWEKQFAKTAI 120

|||||

QY 594 VAIFILSVASKADAVDALKTCTCLLKECRLELAKCISNPACAANVACLQTCNNRPDETEC 653

|||||

Db 121 VAIFILSVASKADAVDALKTCTCLLKECRLELAKCISNPACAANVACLQTCNNRPDETEC 180

|||||

QY 654 QIKCGDLFNSVVDNFENECANVRKCVPRKSDVGFPPDPVSVLVQKFDMDKDFSGKWEIT 713

|||||

Db 181 QIKCGDLFNSVVDNFENECANVRKCVPRKSDVGFPPDPVSVLVQKFDMDKDFSGKWEIT 240

|||||

QY 714 RGLNPTFAFDCQLHEFHTEENKLVGNLSWRTRTPDGGFFTRSAVQKVPQDPKYPGILYN 773

|||||

Db 241 RGLNPTFAFDCQLHEFHTEENKLVGNLSWRTRTPDGGFFTRSAVQKVPQDPKYPGILYN 300

|||||

QY 774 HDNEYLLYODDWYILSSKVENSPEDYIFVYKGRNDADWDGSGSVLYTRSAVLPSIITE 833

|||||

Db 301 HDNEYLLYQDDWYLSKVENSPEDYIFVYKGRNDAMDWDGYSVLYTRSAVLPESIPIE 360
QY 834 LQTAAGKVGDFNTFTKTNTCTGPEPLVERLEKKVEEGERTTIKREVEIEEVEKVRDK 893
Db 361 LQTAAGKVGDFNTFTKTNTCTGPEPLVERLEKKVEEGERTTIKREVEIEEVEKVRDK 420
QY 894 EVTLFSLKFGFKELQDEENFLRELSEKEMDVLGDKMEATEVEKFLGRALPIRKL 950
Db 421 EVTLFSLKFGFKELQDEENFLRELSEKEMDVLGDKMEATEVEKFLGRALPIRKL 477
RESULT 2
T00708
violaxanthin de-epoxidase homolog F22013.3 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999
C:Accession: T00708
R:Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Con-
eologis, A.; Ecker, J.R.
A:Description: Genomic sequence for Arabidopsis thaliana BAC F22013.
A:Reference number: Z14200
A:Accession: T00708
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-462 <SHI>
A:Cross-references: EMBL:AC003981; NID:g3063438; PID:g3063441; GSPDB:GN00059; ATSP:F2201
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: ATSP.F22013.3
A:Map position: 1
A:Introns: 72/3; 128/2; 160/3; 292/2

Query Match 32.4%; Score 2430; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 1,1e-134;
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 951 MAVATHCFTSPCHDRIRFFSSDDGIGRLGITRKRINGTFLKLPQISADLRTTGGRSS 1010
Db 1 MAVATHCFTSPCHDRIRFFSSDDGIGRLGITRKRINGTFLKLPQISADLRTTGGRSS 60
QY 1011 RPLSAFSGSKGIFDIPLPSKNEKELTAPLLKLVGLACAFILVPSADAVDAKTC 1070
Db 61 RPLSAFSGSKGIFDIPLPSKNEKELTAPLLKLVGLACAFILVPSADAVDAKTC 120
QY 1071 ACLKGCRIELAKCIANPACAANVACLOTNNRPDETCQIKGDLFENSVDDEFNECAV 1130
Db 121 ACLKGCRIELAKCIANPACAANVACLOTNNRPDETCQIKGDLFENSVDDEFNECAV 180
QY 1131 SRKCVPRKSDLGEPFAPDPFSLVQNFNISDFNGKWIITSGLNPTDFDCQLHEFHTEG 1190
Db 181 SRKCVPRKSDLGEPFAPDPFSLVQNFNISDFNGKWIITSGLNPTDFDCQLHEFHTEG 240
QY 1191 DNKLGVNISRITKLDGSGFFTRSAVKQFVQDPNQPGVLYNHDNHYLDWILSKIE 1250
Db 241 DNKLGVNISRITKLDGSGFFTRSAVKQFVQDPNQPGVLYNHDNHYLDWILSKIE 300
QY 1251 NKPEYIFVYVYGRNDAMDYGGAIVYTRSVLPNSIIPLEKAASIGRDFSTFTRTDN 1310
Db 301 NKPEYIFVYVYGRNDAMDYGGAIVYTRSVLPNSIIPLEKAASIGRDFSTFTRTDN 360
QY 1311 TCGPEPALVERIEKTVEGRIIVKEVEEIEEVEEVEKVGRTMTLFORLAEGFNELK 1370
Db 361 TCGPEPALVERIEKTVEGRIIVKEVEEIEEVEEVEKVGRTMTLFORLAEGFNELK 420
QY 1371 QDEENFVRELSEKMEFLDBRIKMEASEVEKFLGKALPIRKR 1412
Db 421 QDEENFVRELSEKMEFLDBRIKMEASEVEKFLGKALPIRKR 462

RESULT 3

A84606

hypothetical protein At2g21860 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84606
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: A84606
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-522 <STO>
A:Cross-references: GB:AE002093; NID:g4417279; PIDN:AAD20404.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g21860
A:Map position: 2

Query Match 2.8%; Score 210; DB 2; Length 522;
Best Local Similarity 21.1%; Pred. No. 8.3e-05;
Matches 122; Conservative 77; Mismatches 186; Indels 194; Gaps 30;

QY 375 PEPLVERLEKTAEGEKLLIKEAVEIEEVEEVEKVRDEMTLFORLLEGFKELQODE 434
Db 68 PEPPV--KLIALVKGKGEVSPLK--STSEVEMLMHTARKLWVD-----EGYEMLVFDD 116
QY 435 ENFVRELSEKKEITLNEIQMEATEVEKFLGRALPIRKLALAPHSFLANHETIKYVYG 494
Db 117 E-----IVSSNDORAMN--LTQELNQTDLVVVA-----VNNSESV----- 150
QY 495 SKLPGHKRFESWGWDYFGSIVVAKICSSRRIPRYPRKSPRICCGLDGRGLQFSGHKHNL 554
Db 151 -----NWIQTNSKNVKNMIC-----FESSNIMNRLGG----- 178
QY 555 SPAHSINONVPGKSGCKFPKDVALLV-----WEKMG-----QFAKTAIVAFITLSV--- 601
Db 179 TDVGSVND-----KEVTEVKTGCDAWERRNSDDIRFCLLIVINAYIRPVPL 227
QY 602 ---ASKADAVDALKTCTCLLKECKLELAKCISNPACAANVACLOTNNRPDETCQIKCG 658
Db 228 QNLRSK-----GFSTLSGMWKNCGPQILNCLLDPNCRKALQCLNQS--PVDQVCYRCI 280
QY 659 DLFENSVDDEFNECAVSRKCVPRKSDVGDFPDPSVLVQKFDKDPGSK----- 709
Db 281 ASYEGYFEAFSLCVLQKHNCLELDKTIPEKPYPP-----MTSFRKELCHDTAED 332
QY 710 -----WFTTRGLNPTDFADFCQLHEFH-----TEENKLVGN 740
Db 333 LFGVGLGELEWSRWVAGONPAYDOFPQYQOLFYRGKGSFWEYFPVQVRLTEELV-- 390
QY 741 LSWRLTPDGGFFTRSAVKQFVQDPKYP-----GILYNH--DNEYLLYQDDWYLSKVE 793
Db 391 --WRRR-----RYSVKR-----GKIPATFRFSVLDNGVVSNEF-----WTIVD--VS 428
QY 794 NSPEDYIFVYVYGRNDAMDYGGSVLYTRSAVLP--ESIPELQTAAGKVG--RDNFTFIKT 851
Db 429 DDLWGLFHYHGAARVAGQSYTGAVLVTPDGSYPAEKDKERLQSALEKCGIKELWELF-AV 487
QY 852 DNTCGPEPLVERLEKKVEEGER-----TIKEVEEIEE 885
Db 488 DNCSCENPPL-----GIQGSRLHSRISIIIEEPDSEK 520

RESULT 4

T28677

rhoptry protein - Plasmodium yoelii

C:Species: Plasmodium yoelii

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000

C:Accession: T28677; C45521

R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.

Mol. Biochem. Parasitol. 65, 171-177, 1994

Db 457 IKDFE---KIQDLRACLSSNELKEKESALIDKDDQELNNLRQIKQKKVSESTQSSL 513
QY 277 QTFVQDDPLGALYNHIDNEFLHYQDDWYILSSQIENKPDYIFVYVYGRNDAMDGYGGSV 336
Db 514 QSLORD---ILAEKKKHVEYSQNLAKGELQTE----- 544
QY 337 IYTRSPITLPESTIIPNLOAAKSVGRDFNNFTITDSCGPEPLVERLEKTAEEGEKLLIK 396
Db 545 -----ISNEHLSQSLSIAAEKAAVATNNE-----LSEKNSIQOT 581
QY 397 EAVEIEEVEKEVEKVRDTEM-----TIFORLLGFKELQOEDENFVRELSKEKEILN 450
Db 582 LCNAFOEKLAQSVMLKNEQNFSSLDTSFKKLNEHQELNNHQTITKQL-KDTSSKLIQ 640
QY 451 ELQMEATEVEK-----LFGRALPIRKLRLMALAPH-----SNFLANHETIKYVGS 495
Db 641 QLOLERANFEQKESLTSDENNDLRTKLKLEESNKSILKQEDVDLSLEKNTQTLK----- 695
QY 496 KLPGRKFSWGWEDYFGSIWAKICSSRRIPRYPRKSPRICGLDSRGLQLF---SHGKH 552
Db 696 -----ED-----LRKSEALRFSK-----LEAKNLREVIDNLKRGH 726
QY 553 NLSPA-----HSI-----NONVPKONGCKPPKQDVALMVWEKVGQFAKTAIVAFILSVA 602
Db 727 ETLEQRNDLHSSLSDAKNTNAILSELSTKSSDV-----KRLTANVETLTQD 774
QY 603 SKA-----DAVDALKTCITCLKECRLELAKCISNPACAAVACLOTNNRPDTECOIK 656
Db 775 SKAMQSTSLVNSYQGISNLYHELDRDHVN-----MQSONTLLESESKLK 821
QY 657 --CGDLFEN--SVVDEF-----NECAVRKKCVPRK-----SDVGDFPVPDP 694
Db 822 TDCENLTQONMTLIDNVOKLMMHKHVNQESKVELKYNKLSLDLKNLRSLNVAISDND 881
QY 695 SVLVQKDM-KDFSKWFTITGLNPTFPDQCLHEFTEE-----NKLVGNLWRIR 746
Db 882 QILTQALAEKSNYDSEQASGLKSLAEAKQLLHTENELHILRLDKLTGKLGK----- 937
QY 747 TPDGFFTRSAVQKVPQPKYPGILYNHIDNEYLLYQDDWYILSSKVENSPEDYIFVYK 806
Db 938 -----IEESKSSDL-----GKKLTARQEB--ISNLKEENMSQSAITSVKS 976
QY 807 RNDAMDGYGGSVLYTRSAVLPESIIPELOATAQKVRGDFNTFIKTDNTCGPEPLVERLE 866
Db 977 KLDE-----TLKSCKL-EADIEHLKNKVSSEVERNALLASN-----BELM 1017
QY 867 KVEE-EGERTIKEVEEIEEVEKVRKEVTLFSKL-----FGF-----KELQD 911
Db 1018 DDLKNNGE-----NIASLOTETEKRAENDDLQSKLSVSVSEYENLLLISSQTNKSLE-D 1071
QY 912 EENFLRELSKEEMDVLGKMEATEVEKL---FGR-----ALPIRKLMAVATHC 957
Db 1072 KTNQLUKYTEKNVQKLLDEKQDNVNELEELTSKYGLKEENNAQIKDELLALRK----- 1124
QY 958 FTSPCHDRIRFFSSD---DGIGRL-----GITRKRINGTELLKLLPIIOS--AD 1001
Db 1125 -SKQHDLCANFVDDLEKSDALQELTNEKNELIVLSIQSNNEALVEERSDLANLSD 1183
QY 1002 LRTTGGRRSLPSAPRSGFSGIFDIPLPSKNELKELTAPLLKLVGLACAFIIVPSA 1061
Db 1184 MKKSLSDSDNIVSIRSD-----KQDLSLSTQYSEVCDRDLDSLKGC-----LYRVN 1206
QY 1062 DAVDALKTCALLKGRLELAKCIANPACAAVACLOTNNRPDTECOIKGDLFENS 1121
Db 1207 DELDTLK-----KODLSLSTQYSEVCDRDLDSLKGC----- 1240
QY 1122 VDEFNECAVS-RKCVPRKSDLGEPPADPSLVON--FNISDFNGKWIIT-SGLNPTFD 1177
Db 1241 BESFNKYAVSLRELCTKSEIDV-----PVSILDDNFVFNAGNSELRLTVLSLENYLD 1295
QY 1178 AFDQCLHEFTEGDNKLVGNISWRIKTLDSGFFTRSAVQKVPQNPQGVLYNHIDNEYLH 1237
Db 1296 AFN-QVNFKKMELDN-----RLTTTDAEF-----TKVVADL-----EKLQH 1330

QY 1238 YQDDWYILSSKIENKPEDYIFVYVYGRNDAMDGYGGAVVYTRSSVLPNSIPIELEKAAS 1297
Db 1331 EHDDWLI-----QRG-----DLERALKD 1348
QY 1298 IGRDFTFIRTDNTCGPEPALVERIEKTVEGEGRIVKVEEIEEVE----- 1345
Db 1349 SEKNF-----LRKEAEMTENIH-SLEGKEETKKEIAELSSRLQDNLATNKLKQ 1398
QY 1346 ----KEV-----EKVGRTEMTLFORLAEGFNEKLQDBENFVRELSKEMEPL--DEIKMEA 1395
Db 1399 LDHLNQELRKEDVLKESKESLIISLEESLNQKQESSLLD--AKNELEHMLDDTSRKNS 1456
QY 1396 SEVEKL 1401
Db 1457 SLMEKI 1462
RESULT 7
T30822
Impl protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 07-Dec-1999
C:Accession: T30822
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christiansen, G.
Infect. Immun. 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis PG21 deletion mutants by cultivation in the
A:Reference number: Z18884; MUID:95369882
A:Accession: T30822
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1365 <JEN>
A:Cross-references: EMBL:U021962; NID:g790243; PID:g790244; PIDN:AAA81013.1
C:Genetics:
A:Gene: Impl
A:Genetic code: SGC3

Query Match 2.1%; Score 156; DB 2; Length 1365;
Best local similarity 16.8%; Pred. No. 0.46;
Matches 209; Conservative 187; Mismatches 420; Indels 430; Gaps 52;

QY 272 TRSAVOTFVQDDPLGALYNHIDNEFLHYQDDWYILSSQIENKPDYIFVYVYGRNDAMD 331
Db 282 TRNQIOEFINT-----NKNAP--NYSE-----LISQTSKRDD-----SKNSVTS 319
QY 332 YGGSVIYTRSPITLPESTI--PNLOKA-----AKSVGRDFNNFTITDSCGPEPLVERLE 385
Db 320 SNKSDIESANTELKQALAKANADKQVADNLAKSIKQLNNVSNANT-----LSA 369
QY 386 TABEGEKLLIKEAVEIEEVEKEVEKVRDTEMPTLFORLLEGF---KELQOEDENFVREL 441
Db 370 KLTDKONTIQQAQTEKEVQKADQAIKSNNTASMQSAKSSLDKVAEITKKLETF---- 425
QY 442 SKEEKELLINELQMEATEVEKLFGRALPIRKLRLMALPHSNFLANHETIKYVGSKLPGHK 501
Db 426 NKDEAKFNELKOTRQIOEF-----ADKVAQ---DNLA 446
QY 502 RFSWGWEDYFGSIWAKICSSRRIPRYPRKSPRICGLDSRGLQLFSGHKNLSPAHS-I 560
Db 447 -----INTNKNPNYSELISQTSKRDSKNSVTSNKSIESANTSL 489
QY 561 NONVPKONGCKPPKQDVALMVWEKVGQFAKTAIVAFILSVASKADADALKTCTCLLKE 620
Db 490 KQALAKAN-----ADKVAQ---DNLA 509
QY 621 CRLELAKCISNPACAAVACLOTNNRPDTECOIKGDLFENSVDENECASRKKCV 680
Db 510 IKQLNLSVSN-----ANTLSAKLTDKDN--TIOQAKTE 541
QY 681 PRKSDVGDFFVPDPSPVLVQKFDKMGDFSGKWFITRGLNPTFDA-----FDCQLHEFTEENK 736
Db 542 LEKE-----VOKAQAIKSNNTASMQSAKSSLDKVAEITKKLETFNKKDEA 588

C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

```

QY 737 LVGNLSWRIPTDGGFTSAVOKFVQDPKPGILYNHNDNEYLLODDWYILSSKVENSP 796
Db 589 KFNELKO-----TRNQIEFTINTK-----NNPN-----YSELISQITSKRDS-- 626
QY 797 EDYIFVYVYKGRNDWDGCGSVLYTRSAVLPEIIP-----ELQTAQKVGGRDENTFIK 850
Db 627 -----KNSVTDSSNKSDESANTELKQALANADKQVADNLAQKSIKQLNNSVS 676
QY 851 TDNTCGPEPLVERLEKKVEGERTIIEVEEIEVEK-----VDRKEVTLF 898
Db 677 NANP-----LSAKLTDKDNTIQQAQTELEKIQANQANQAIKSNNTASMQSASSLD 726
QY 899 SKLPEGEKELQ--RDENFLRELSKEEMDVGLKMEATEVEKLFGRALPIKLMAVAT 955
Db 727 AKVAETIKKLETKENKDEAFNELKQTRNOI-----QEFINTNK----- 765
QY 956 HCFSTPCHDRIRFFSDDGIGRIGITRKIRINGFTLLKILPPIOSAD-----LRTTGGRS 1009
Db 766 ---NNPNYSEL-----ISQTSKRDSKNSVTDSSNKSDESANTELKQALNTAKAK 814
QY 1010 S-----RPLSAPRSFGSKGIFDIVL-----PSKNEL-KELTAPLALLKLVG 1049
Db 815 SSIDNEURPL---KNDLQSKIEFGPIRNTNFWISKSLETTNKLAELT----- 862
QY 1050 VLCAFLIVPSADAV-----DAKTCACLLKGRICLAKCIANPACAANVACIQTNC 1101
Db 863 -----KADAIKNPNSSQALKDSSQVQKGLGNELTKITE-----EFGKVETKN 907
QY 1102 NRPDETCQJLKGDLFNSVVDENECASVKCVPRKSDLGFEFPAPD-----PSVIVQ 1155
Db 908 SNIGYRLFKLAQAFQFNNSDQKLNKAWEEKOTLLSKKQKLNQSTKDYLTQSTEMSTQ 967
QY 1156 NFNTSD--FNGKWITSLGNTFTDAFCQLEHFEHTEGDNKLVGNISWRIKTLDSGFFTR- 1212
Db 968 ESTKKVIVNIOAHIRNLN-----SOYRLEAD-KLIAN-----MKRGYGDV 1009
QY 1213 --SAVKFVQDPNPGVLYNHNDNEYLHYQDW-----YILSSKIENKPEDYIFVYVYGRN 1265
Db 1010 GIESLQKW-QDLMDSDSVLSVDDSDS-----LKDDFNKALRVLGDTYKPNPVSSFWINKRNS 1064
QY 1266 -DAWDGCGGAVVYTRSSVPSNSI-----IPELKAASIGR----- 1300
Db 1065 IENTQNRLNLLVRENELIDKADLDKRAEKTIKFVDENINSLDQAKRLKOEILNKN 1124
QY 1301 DFSTFIR-----TDNTCGPEPALVERIEKTVVEGERIIVKE-----VEEIEEVEKEV 1348
Db 1125 DLSNFTLNHQKNQFTAKDITPKISLLENKLEINQYLLPIIKEKAVSKISEIEKN-KKEL 1183
QY 1349 EKVGRTEMTLQRLAEGFNLKODEENFVRELSKEEMEFLEIKME 1394
Db 1184 EDIIRSNFYLWE-----KVEINKYIISLTNKNQVLEKRSNINFE 1220

RESULT 8
H69378
Conserved hypothetical protein AF1032 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Feb-2001
C:Accession: H69378
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 384-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, S.
; Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343
A:Accession: H69378
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-886 <KLE>
A:Cross-references: GB:AE001032; GB:AE000782; NID:g2689355; PIDN:AAB90211.1; PID:g264958

```

Query Match 2.18; Score 155.5; DB 2; Length 886;

Best Local Similarity 20.6%; Pred. No. 0.27; 221; Indels 221; Gaps 27;

Matches 137; Conservative 85; Mismatches 221; Indels 221; Gaps 27;

```

QY 797 EDYIFVYVYKGRNDWDGCGSV--LYTRSAVLPEIIPQLQTAQKVGGRDENTFIKTD-N 853
Db 152 EDY-----ENAWKNLGAIVRMLEKEREKLEKLFQEQEIKRQKEK-----KAEIE 197
QY 854 TCGPPPLVERLEKKVEGERTI---KVEEIEEVEKVRDKEVTLFSKLFEKFKELQR 910
Db 198 RISEIKIESLREKLSSEVRNLESRLKELEHKSRLSEKQESSVLQEV-----R 249
QY 911 DEENFLRELSKEEMDV---LDGLKMEATEVEKLFGRALPIKLMAVATHCTSPCHDRIR 967
Db 250 GLEKRLREKQLKEVVRIEDLEKKAKEVKEL-----KPAERY- 289
QY 968 FFSDDGIGRIGITRKIRINGFTLLKILPPIQSA--DLRTTGGRSRSLPSAFRSFGSGIF 1025
Db 290 -----SILEKLLSEINOALRDVREKREGDLTREAAAGIQAOQKKAEE 329
QY 1026 DIVPLPSKNELKELTAPL-----LLKLVGLACAFLLIVPSADAVDAKTC 1070
Db 330 D-----NSKLEETIKRIEELERELERPEKSHRLL-----TLKPKMDRMQGIK-- 372
QY 1071 ACLKGCRIELAKCIANPACAANVACIQTNNRPDETETECQIKCGDLFNSVYDFNECAV 1130
Db 373 -----AKLEKNLTPDKVE---KMYDILLSKAKEE-KEITE 404
QY 1131 SKKCVPRKSDLGEPAPDPVSVLVONFNSDFNGKWIYITSLGNFTFADFQCLHEFHTEG 1190
Db 405 KLLKLIARKSSILKTRGAOLKKAEE-----LKSAERTCPVCGRELDEEHRKN 451
QY 1191 -----DNKLVGNISWRIKTLDSGFFTRSAVOKFVQDPNPGVLYN-- 1230
Db 452 IMAEYTRMKRIAEELAKADIEKKLEKLEKVEKALEKQETVLKYRQWVDELKALENEL 511
QY 1231 --HDNEYLHYQDWYILSSKIENKPEDYIFVYVYGRNDWDGCGGAVVYTRSSVPSNSII 1288
Db 512 SSHDAE-----KLSAESE-----YRKVKERLDGLRQGOQKILLSSA---SRI 550
QY 1289 PELEKAASIGRDFSTFIRTDNTCGPEPALVERIEKTVVEGERIIVKEVEIEEVEKEV 1348
Db 551 KEKLSLREI-----EAL-KNVESEGLHRLKIREGFSLELEREV 593
QY 1349 EKVGRTEMTLQRLAEGFN---ELKQDEENFVRELSKEEMEFLEIKMEASEVEKLFQKA 1405
Db 594 -----QSLRPFYNNKWLKELKDAESLELKRRE-KLEDEISRAIAKLEANGKA 641
QY 1406 LPDR 1409
Db 642 EDIR 645

```

RESULT 9

A42771
 reticulocyte-binding protein 1 - Plasmodium vivax
 C:Species: Plasmodium vivax
 C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995
 C:Accession: A42771
 R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
 Cell 69, 1213-1226, 1992
 A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
 A:Reference number: A42771; MUID:92315338
 A:Accession: A42771
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2829 <GAL>
 A:Experimental source: Belem strain, merozoites
 A>Note: sequence extracted from NCBI backbone (NCBIN:108114, NCBIPI:108115)


```

QY 847 TFIKTDNTCGPEPLVERLEKKVEGERTIIKEVEEIEEEVEKYR-----DKEVT 896
DB 414 LEIE-----TIKLNKEEIEELEKKNKAHTVEVTNLTKEIK 449
QY 897 LFSKLFEGFKELORDE----ENFLRELSKEEMVDLDGLKMEATEVEKLFGRALPIRKIMA 952
DB 450 LLEKTEDEAKGHNELNQLSKLNKEK---DNIKNENTEL-----490
QY 953 VATHCFTSPCHDRIRTFSSDDGIGRLGITRKRINGTFLKILPIQISADLRTTGGSSRP 1012
DB 491 -----NDKISSNSE-----VN-----ILNKDKQTLG-----512
QY 1013 LSAPRSFGSKGIFDVLPLPSKNELKELTAPLLKLVGLVLAFLVPSADAVDAKTCAC 1072
DB 513 -----NDIKTLN-----DLINLNK-----527
QY 1073 LLKGCRIELAKCIANPACAAVACLOTNNRPDE-----TECOIKGDLFENSVD 1123
DB 528 -----EINTSDKNMKWKEDELAWLNEEMGKC-----VVID 558
QY 1124 EF-----NECAVSRKKCVPRK--SDLGEFFPADPDSVLVQNFNIDFPNGKWIYTSGLNPTF 1176
DB 559 EIEKKYKNEIFMLEKLEKENYADLND---EISILRNSIYVKE---KEFI-----603
QY 1177 DAFDCOLHEFHTEGDKLVGNISWRIKTLDSGFFTSRQVOK---FVQDPNQPGLVYNHDN 1233
DB 604 -----EMKEFY---ENKI-----NLFKNFEKKNIYENELSLRLKYDNEQ 642
QY 1234 BYLHYQDDWYILSSKIENKPEYIFVYVYGRNDAMDYGAVVYTRSSVLPNSIIPLEK 1293
DB 643 GLIKQIDELNIQKLTKEK---YLQLY---NDN-----MHMFRSICKIDMPYSEN 687
QY 1294 AAKSIGRDFST-FI--RTDNTCGPEPALVERIEKTVVEGERIIVKEVEEIEEVEKEVEK 1350
DB 688 IKGSDLVDFVYAYIKRDESSDANPDTHK--EMVAELEKRAHAAIYVAELEKHKEIAK 745
QY 1351 VGRTEMTLQRLAEGFNE---LKODENFVRELSKEEMFLEIDIKMEASEVEKLF 1403
DB 746 LGEHKEVVLRLGQHKETEILLEKHKVDVTKLGEQHKENIKLEEHKDVVTKLG 802

RESULT 12
A59286
myosin heavy chain beta chain, cardiac - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C:Accession: A59286
R:Ko, Y.L.
submitted to GenBank, October 1996
A:Reference number: A59286
A:Accession: A59286
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1935 <R0>
A:Cross-references: GB:U75316; NID:g1698894; PIDN:AA37320.1; PID:g1698895
A:Experimental source: strain domestica
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: cardiac muscle; heart
F:88-766/Domain: myosin motor domain homology <M0>

Query Match 2.0%; Score 148.5; DB 2; Length 1935;
Best Local Similarity 16.7%; Pred. No. 2;
Matches 248; Conservative 220; Mismatches 555; Indels 459; Gaps 57;

QY 166 PDETECOIKGDLFENSVDQFNCAVSRKKCVPRKSDVGEFFVPDRN-----AVV 216
DB 375 PDGTE-----EADKSAYLMGLNSADLLKGLCHPRVKVGNVYVTKGONVQVYATGALA 428
QY 217 QNFENKDFSGKWIYTSGLNPT-----FADFDCQLHE---FHMENDKL 255
DB 429 KAVYERMFN---WMVTR-INTTLETKOPRQYFVGLVDIAGFEIFDFNSFEOLCINFNTNEKL 485

```

```

QY 256 -----VGNLTWRIKTLGGFTTSRAVOTF-----V 280
DB 486 QQPFNHMFVLEOEYKKEGIEWEF--IDFGMDLQACIDLIEKPMGIMSILEECMFPPKA 543
QY 281 QDPDLPCALY-NHIDNEFLHVQDDWYILSSQIENKPD-DVIFVYVYGRND---ANDGYGG 334
DB 544 TDMTFPAKLYDNHILGKSNFQK-----PRNIKGRPPAHFALIHVAGTVDYNIIGW-----593
QY 335 SVIYTRSPILPESIIPLNQAAA--KSVGRDFNNFITDNSCGPEPLVERLEKTAEEGKL 393
DB 594 --LQKNKDPNETVVDLYKKSSLLKLSNLFANY-----AGADTP-VEKGKCAKKGSF 644
QY 394 LIKEAVEIEEVEKEVEKYRVDRTMTLQRLLEGFKELOODEENFVRELSKEKE-----447
DB 645 QTVSALH-RENLMKLTNLRST-----HPHFVRCIIPNTPSKPGVID 685
QY 448 ---ILNELOMEAT-EVEKLFGRALP-----IRKLRLMALAPHSNFLANHETIKY 492
DB 686 NPLVMHQLRCNGVLEGIRICRKGFPNRIYLGDFRQRYRILNPAAPGQFIDSRKGAEL 745
QY 493 VGSKLPGHKRFSGWEDYFGSIIVAKICSSRRIPRYFRKSPRICCGLDLGRGLQFSGHKG 552
DB 746 LGSLLDIDHNOYKFGHTKVFVKAGLLGLEEMRDERLSRIITRI--QAQSRGVLSEMEFKK 803
QY 553 NLSPAHS---INONVPKNGSGCKFPKDVALMVWKGQPAKTAIYVAIFLTSVASKADAVD 609
DB 804 LLERRDSLIIOWNIIR-----AFMSVKNW-PWMKLYFKIKPLLESAETEKEMA 850
QY 610 AKTCTCLLKEC-----RLELAKCISNPACAAVACLOTNNRPDETECOIKGDLF 662
DB 851 TKEEEGRLKEALEKSEARKELEKMWLSLQEKNDLQLOVQAEQNLSDSEERCDQLIK 910
QY 663 NSV-----VDEFNECAVSRKK-----CVPRKSDVGDVFPVDPSPVLVQKFD 702
DB 911 NKIQLEAKVKEMTERLEDEEMNAELTAKKNVEDECSSELKRDIDDLLETLAKVEKEK--968
QY 703 MKDFSGKWIITRGLNPTFAFDCQLHEFTEENKLVNLSWRIPTPDGFGFFTSRAVQRFV 762
DB 969 -----HATENK-VKNUTEEMAGLD-----ETIAKIT 993
QY 763 QDPKPGIILYHNDNEYLQDDWYILSSKVENSPEDYIFVYVYKGRNDAMDYGGSVLYTR 822
DB 994 KEKKAQEAHQALDDLOAEEDKVNLTAKVKLEQHV-----DDLAGSLEQEK 1042
QY 823 SA-----VLPEIIPELQTAOKVG-----RDF-----NTFTKTONTQCP 857
DB 1043 KYRMDLERAKRKLEGLDKLTQESIMDLNDKQOLDERLAKKDFELNALNARIEDEQALGS 1102
QY 858 E-----PPLVERLEKKVEGERT-----LIKEVEEIEEVEK-----889
DB 1103 QLOKKLKELOARIEELEEL-EAERTARAKVEKLSRSDLSRELEEISERLEEAGGATSVOI 1161
QY 890 -----VRDKVETLFSKLFEGFKELQORDEENFLREL 919
DB 1162 EMKKREAEFOKMRDRLEATLQHEATAAALRKHADSVAEELGEQIDNLQRYKQLEKEK 1221
QY 920 SKEEMDVLGLKMEATEVEKLFGRALPIRKLMVATHCTSPCHDRIRTFSSDDGIGRLG 979
DB 1222 SEFKLELDD---VTSNMEOIIRKAKANLEK-----CRT--LEDQMNEHRS-----RAE 1264
QY 980 ITRKRINGTFLKILPPIQISADLRTTGGSSRPL---SAFRSGFSKGIQFIDVPLPSKNEL 1036
DB 1265 ETQORSVNDL-----TSORAKLQTEENGELSRLQDEKEALISQLTRG-----KLTYTQOL 1312
QY 1037 KELTAPLLKLVGLVLAFLVPSADAVDAKTCACLLKGCRIELAKCIANPACAAVAC 1096
DB 1313 EDLKRQLEEEVAKNALAHALQASARHAADLLREQYEEETETKAELQRLSK--ANSEVAQ 1370
QY 1097 LOTCNRNPDETECOIKGDLFENSVDENCAVSRKKCVPRKSDLGEFFPADPSPVLVQ 1156
DB 1371 WRT-----KYETDAIQRTTEELEAKKLAQRLQDAEE-----1402

```

QY 1157 FNISDFNGKWIITSGNPTDFADFCOLHEFHETGDNKLVGNISWRIKTLDSGFFTRSAVQ 1216
 Db 1403 -----AVEAVNAKCSLEKTKHRLONEIEDLMDV-----VERSNAA 1438
 QY 1217 KFVODPNQGVLYHNHONEYLHYODDYWILLSSKTKENKPEYIFVYVGRNDAMDGGYGVAV 1276
 Db 1439 AALADKKQ-----RNFDKILAEWKQYKESQSELESQKE-----ARSL 1477
 QY 1277 YTRSSVLPNSIIPELE-----KAAKSIGRDFSTFTDNTCGPEPALVERIEKTVVEE 1330
 Db 1478 STEFLKLNAYEESLEHLETSKREKNLQAEISDLTEQLGSSGKTHLEKVRKKQLEAK 1537
 QY 1331 RIIVKEVEEIEEVEKEVEKVRTEMTLQ-----RLAEGFNELQDEENFVRELS-- 1381
 Db 1538 LELQSALEESSEASLEHEGKILRAQLAEFNQIAEMERKLAKEDEEMQAKRNLHRLVVDLS 1597
 QY 1382 -----KEEME-FLDEIRKEASEVEKLFGKA 1405
 Db 1598 QTSIDAETRSRNEALRVKKKMEGDLNEMETQLSHANMAEA 1639
 RESULT 13
 A37102
 myosin beta heavy chain, cardiac and skeletal muscle - human
 N:Contains: myosin ATPase (EC 3.6.1.32)
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Jan-2001
 C:Accession: A37102; S12733; A94224; B28908; A24997; A27858; S12458; S09331; S02
 R:Jaenicke, T.; Diederich, K.W.; Haas, W.; Schleich, J.; Lichter, P.; Pfordt, M.; Bach,
 Genomics 8, 194-206, 1990
 A:Title: The complete sequence of the human beta-myosin heavy chain gene and a comparati
 A:Reference number: A37102; MUID:91065634
 A:Accession: A37102
 A:Molecule type: DNA
 A:Residues: 1-1935 <JAE>
 A:Cross-references: GB:M57965; GB:M30603; NID:g179507; PIDN:AAA51837.1; PID:g179508; GB:
 R:Liew, C.C.; Sole, M.J.; Yamauchi-Takahara, K.; Kellam, B.; Anderson, D.H.; Lin, L.; Li
 Nucleic Acids Res. 18, 3647-3651, 1990
 A:Title: Complete sequence and organization of the human cardiac beta-myosin heavy chain
 A:Reference number: S12733; MUID:90301496
 A:Accession: S12733
 A:Molecule type: DNA
 A:Residues: 1-106, 'E', 108-671, 'LYH', 675-857, 'A', 859-941, 'NV', 944-1123, 'A', 1125-1158, 'C',
 A:Cross-references: EMBL:X52889; NID:g29726; PIDN:CAA37068.1; PID:g29727
 R:Yamauchi-Takahara, K.; Sole, M.J.; Liew, J.; Ing, D.; Liew, C.C.
 Proc. Natl. Acad. Sci. U.S.A. 86, 3504-3508, 1989
 A:Title: Characterization of human cardiac myosin heavy chain genes.
 A:Reference number: A94224; MUID:89264452
 A:Accession: A94224
 A:Molecule type: DNA
 A:Residues: 1-87, 'Q', 89-106, 'E', 108-177, 1325-1702, 'DR', 1705-1786, 1788-1803, 'E', 1804-1935
 R:Yamauchi-Takahara, K.; Sole, M.J.; Liew, J.; Ing, D.; Liew, C.C.
 Proc. Natl. Acad. Sci. U.S.A. 86, 7416-7417, 1989
 A:Reference number: A94226
 A:Contents: annotation; erratum
 R:Kurabayashi, M.; Tsuchimochi, H.; Komuro, I.; Takaku, F.; Yazaki, Y.
 J. Clin. Invest. 82, 524-531, 1988
 A:Title: Molecular cloning and characterization of human cardiac alpha- and beta-form my
 human atrium.
 A:Reference number: A92770; MUID:88299163
 A:Accession: B28908
 A:Molecule type: mRNA
 A:Residues: 1412-1518, 'R', 1520-1574, 'NV', 1577-1935 <KUR>
 A:Cross-references: GB:M21665
 A:Note: the authors translated the codon AGC for residue 108 as Arg
 R:Lichter, P.; Umeda, P.K.; Levin, J.E.; Vosberg, H.P.
 Eur. J. Biochem. 160, 419-426, 1986
 A:Title: Partial characterization of the human beta-myosin heavy-chain gene which is exp
 A:Reference number: A24997; MUID:87030293
 A:Accession: A24997
 A:Molecule type: DNA
 A:Residues: 682-721; 975-1112; 1854-1935 <LIC>
 A:Cross-references: GB:X04627

R:Saez, L.J.; Gianola, K.M.; McNally, E.M.; Fedhali, R.; Eddy, R.; Shows, T.B.; Leinw
 Nucleic Acids Res. 15, 5443-5459, 1987
 A:Title: Human cardiac myosin heavy chain genes and their linkage in the genome.
 A:Reference number: A93669; MUID:87260010
 A:Accession: A27858
 A:Molecule type: DNA
 A:Residues: 1854-1865, 'A', 1867-1935 <SAE>
 A:Cross-references: GB:X05631; GB:Y00362; NID:g34643; PIDN:CAA29119.1; PID:g34644
 R:Diederich, K.W.; Eisele, I.; Ried, T.; Jaenicke, T.; Lichter, P.; Vosberg, H.P.
 Hum. Genet. 81, 214-220, 1989
 A:Title: Isolation and characterization of the complete human beta-myosin heavy chain
 A:Reference number: 154254; MUID:89154425
 A:Accession: 154254
 A:Status: translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 653-720 <RES>
 A:Cross-references: GB:M27636; NID:g179511; PIDN:AAA79019.1; PID:g601916
 R:Bober, E.
 submitted to the EMBL Data Library, January 1989
 A:Reference number: S12458
 A:Accession: S12458
 A:Molecule type: mRNA
 A:Residues: 785-1076, 'E', 1078-1123, 'A', 1125-1702, 'DE', 1705-1935 <BOB>
 A:Cross-references: EMBL:X51591; NID:g29467; PIDN:CAA35940.1; PID:g29468
 R:Bober, E.; Buchberger-Seidl, A.; Braun, T.; Singh, S.; Goedde, H.W.; Arnold, H.H.
 Eur. J. Biochem. 189, 55-65, 1990
 A:Title: Identification of three developmentally controlled isoforms of human myosin
 A:Reference number: S09331; MUID:90235862
 A:Accession: S09331
 A:Molecule type: mRNA
 A:Residues: 785-830, 'X', 832-900, 'X', 902-970, 'X', 972-1040, 'X', 1042-1076, 'E', 1078-1110,
 '1602-1670, 'X', 1672-1702, 'DE', 1705-1740, 'X', 1742-1810, 'X', 1812-1935 <BOB>
 R:Jandreski, M.A.; Liew, C.C.
 Hum. Genet. 76, 47-53, 1987
 A:Title: Construction of a human ventricular cDNA library and characterization of a b
 A:Reference number: S02229; MUID:87192738
 A:Accession: S02229
 A:Molecule type: mRNA
 A:Residues: 1393-1702, 'DR', 1705-1935 <JAN>
 A:Cross-references: EMBL:X06976; NID:g34860; PIDN:CAA30039.1; PID:g825694
 R:Saez, L.; Leinwand, L.A.
 Nucleic Acids Res. 14, 2951-2969, 1986
 A:Title: Characterization of diverse forms of myosin heavy chain expressed in adult h
 A:Reference number: A93616; MUID:86176776
 A:Accession: B23767
 A:Molecule type: mRNA
 A:Residues: 'LLGVGELASG', 1311-1312, 'G', 1314-1355, 'R', 1357-1358, 'GD', 1361-1438, 'LQ', 14
 A:Note: the first ten codons of the sequence figure show the reverse complementary st
 C:Genetics:
 A:Gene: GDB:MYH7
 A:Cross-references: GDB:120215; OMIM:160760
 A:Map position: 14q12-14q12
 A:Introns: 67/3; 115/3; 168/1; 177/2; 213/3; 244/3; 266/1; 299/1; 333/3; 380/1; 419/3
 24/3; 1390/2; 1451/3; 1507/1; 1548/3; 1651/3; 1719/3; 1761/3; 1853/3; 1885/3; 1930/3
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methyl
 F:88-766/Domain: myosin motor domain homology <MMOT>
 F:178-185/Region: nucleotide-binding motif A (P-loop)
 F:548-585/Region: actin binding #status predicted
 F:655-677/Region: actin binding #status predicted
 F:839-1935/Domain: coiled coil #status predicted <COI>
 F:839-1279/Region: S2
 F:1280-1935/Region: light meromyosin
 F:129/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
 F:184/Binding site: ATP (Lys) #status predicted
 F:695,705/Active site: Cys #status predicted

Query Match 1.9%; Score 146; DB 1; Length 1935;
 Best Local Similarity 16.9%; Pred. No. 2.9;
 Matches 238; Conservative 207; Mismatches 529; Indels 434; Gaps 51;

Qy	203	DVGEPVDDRN	AVVQ-----NFNMKDFSGKWYITSGLNPFTDFAFDQCLHEF	244	
Db	461	DIAGFEIDFDFNS	FQLCINFETNEKLOQFFNHMHFVLEQOEYKKEGTWFTDFGMDLQAC	520	
Qy	249	HMENDKLVG--	-----NLTWRKTLTDCGGFFTRS	AVQTFVQDPLDGLALYNH	293
Db	521	IDLLEKPMGINS	ILEBECMPKATDMTFKAKLPD-----NHL	557	
Qy	294	NEFLHYODDY	YILSSQIENKPD-DYIFVYGRND---	AWDGYGGSVITYTRSP	348
Db	558	GKSANFOK----	PRNKGKPEAHFSLIHYAGIVDYNIGW-----LQKNKQDPLNETV	605	
Qy	349	IPNLQKAA-KSVGRDN	NFTITDNSCGPPEPPLVERLEKTAERGEKLLLKEAVEIEEVEEK	407	
Db	606	VGLYKSSLLTSL	FANY-----AGADAP-IERGKRAKKGSS	FQTVSALH-REN	657
Qy	408	EVEKVRDTEMT	LFQRLLEGFKELQODEENFVRELSKEEKE-----ILNETLOMEAT-	457	
Db	658	LMTNLRT-----	HPHEVRCIIPNETKSPGVMDNPLVMHQLRCNGVL	696	
Qy	458	EVEKLFGRALP-----	IRKLRMALAPHSNPLANHETIKYVYGSKPLPGHKRPSWG	506	
Db	700	EGIRICRKGPN	RILYDGFQRQRYRILNPAAPGOFIDSRKGAERKLLSLDIDHNOYKFG	759	
Qy	507	WEDYFGSIV	VAKICSSRRIPRVFRKSPRCCGLDSRGLQFSGHKHNSPAHS-----INQN	563	
Db	760	HTKVFFKAGL	LGLEWRDERLSRIITRI--QAOSRGVLARMEYKLLERDRSLLVTOVN	817	
Qy	564	VPKNSGCKP	KDVALMWKEWQOFAKTAIVAIFILSVASKADADALKTCTCLKEC--	621	
Db	818	IR-----	APMGVKNW-PWMKLYFKIKPLLKSAEREKEMASMKEEFTRLKEALE	864	
Qy	622	-----RLELAK	ISNPACAANVACIOTCNNRDETECQIKCGDLFENSV-----VDEPNE	671	
Db	865	KSEARRKELEK	MVSLLOEKNDLQJQVQAEQDNLADEERCQLTKNRIQLEAKYKEMNE	924	
Qy	672	-----CAVSRK--	KCVPRKSDVGDFPVPDPSVLVQKPFMDKDFSGKWFITRGL	716	
Db	925	RLEDEEMNAEL	TAKRKLDECESELKRDIDDLTLTAKVEKEK-----	968	
Qy	717	NPTFDADCOL	HEFHTTEENKLVNLSWRIRTPDGGFFTRSAVQKVPQDPKYPGILYNHDN	776	
Db	969	-----HATENK-	VKNLTEEMAGLD-----ETIAKLTBKKAQLEAHQOAL	1007	
Qy	777	EYLLYQDDWY	ILSSKVENSPEDYIFVYKGRNDAMDGYGGSVLYTRSA-----	824	
Db	1008	DDLQABEDK	VNTLTAKVKLEQV-----DDLLEGSLEQEKVMDLERAKRKL	1056	
Qy	825	---VLPE	IIPELOTAQKV-----RDP-----NTFIKTONTWCP	863	
Db	1057	GDULKLOES	IMDLENDKOQDLERLKKOPFELNALNARIEEQALGSQLKKLKELOARIE	1116	
Qy	864	RLKKYVEEGERT	-----IIKEVEIEEVEEK-----	889	
Db	1117	ELEBEL-ESERT	ARAKVEKLRSDLSRELEESERLEEAGGATSVQIEMKKRREAFQKMR	1175	
Qy	890	-----VRQK	VTLSFKLPEGFKELORDEENPLRELSKEEMDVLDGLKME	933	
Db	1176	RDLEEATLQ	EATAAALRRKKHADSVAELGEQIDNQRVQKLEKSEKFLEDD--	V	1231
Qy	934	ATEVEKLFGRALP	IRKLMVATHCTFSPCHDRIRFFSDDGI	GRLGITRKIRNGTFFLLKI	993
Db	1232	TSNWEQII	AKANLEK-----CRT--LEDQMNDEHS-----RAEETQRSVNDL	-----	1273
Qy	994	LPPIQSADL	RTTGGSSRPL-----SAFRSGFSKGI	FDIVPLPFSKNELKBLTAPLLKLKLVG	1050
Db	1274	--TSORAKLO	TENGELSRLDEKEALISOLTRG-----KLTYTQOLEDL	KRQLEEVKAK	1326
Qy	1051	LACAFI	VPASADAVALKTCACLLKGRCLIAKCIANTPACAANVACIOTCNNRDETECO	1110	
Db	1327	NALAHQ	ASRHDCLLRBOYEETEETAKALQRLVLSK--ANSEVAQWRT-----	1373	
Qy	1111	IKCGDLFENS	VDFNECAVSRKKCVPRKSDGPEPPDP	SVLVQNFNISDFNGKWITS	1170

Db 1374 -----KYETDAIQRTSELEBAKKKLAAQLQEAEE-----AVL 1405

QY 1171 GLNPFTAFDCQLHEFTEGDNKLVGNISWRKLTLDGFGFTTSAYQKVFQDPNPGVLYN 1230

Db 1406 AVNAKCSLEKTKHRLQNEITDLMVD-----VERSSAAAAALDKQ-----RN 1448

QY 1231 HDNEYLHYQDDWYLLSSKIENKPEDIFVYGRNDANDWGYGAVVYTRSSVLPNSIPE 1290

Db 1449 FDKILAEWKQKYSESSESSQKE-----ARSLSTELFKLNAYEES 1491

QY 1291 LE-----KAAKSIGRDFSFITDNTCGPEPALVERIEKTVEEGERIIIVVEEIEEEV 1344

Db 1492 LEHLETKRENKNLQOEISDLTLEQSSGKTIHELEKVKQLAEKMEQLSALEAEASL 1551

QY 1345 EKEVEKVGRTMTLFQ-----RLAEGNELKQDEENPVRELS----- 1381

Db 1552 EHEGKILRAQLEFNQIKAEIERKLAEKDEDEQAQRNHLRVVDVSLQTSLEAETRSNEA 1611

QY 1382 --KEEME-FLDEIKMEASEVEKLFGKA 1405

Db 1612 LRVKKMEGDLNEMEIQLSHANMAEA 1639

RESULT 14

S57976

nuclear migration protein NUM1 - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YD8358.06; protein YDR150W

C:Species: Saccharomyces cerevisiae

C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000

C:Accession: S57976; S19052; S17018

R:Murphy, L.; Richards, C.; Harris, D.

submitted to the EMBL Data Library, July 1995

A:Reference number: S57971

A:Accession: S57976

A:Molecule type: DNA

A:Residues: 1-2748 <MUR>

A:Cross-references: EMBL:Z50046; NID:g899393; PID:g899399; MIPS:YDR150W

A:Experimental source: strain AB972

R:Kormanec, J.; Schaaff-Gerstenschlaeger, I.; Zimmermann, F.K.; Perecko, D.

Mol. Gen. Genet. 230, 277-287, 1991

A:Title: Nuclear migration in Saccharomyces cerevisiae is controlled by the

A:Reference number: S19052; MUID:92079907

A:Accession: S19052

A:Molecule type: DNA

A:Residues: 1-1569,'V',1571-1821,'K',1823-1959,'RHL',1963-1970,'RN',1973-20

A:Cross-references: EMBL:X61236; NID:g4071; PIDN:CAA43554.1; PID:g4072

C:Genetics:

A:Gene: SGB:NUM1

A:Cross-references: SGD:S0002557; MIPS:YDR150W

A:Map position: 4R

Query Match 1.9%; Score 146; DB 2; Length 2748;

Best Local Similarity 17.9%; Pred. No. 4.7; Mismatches 581; Indels 486; Gaps

Matches 283; Conservative 232;

Qy 62 SSFSDSHCKDKSQICSIDTSFEETQRFDLKRGMTLLLEKQWRF-----IQLAIVLC 115

Db 748 SAYEDLVKCKENPDV-----EFLKEKSAKLGHTVWSSEYSELQRYKSELEKEVQPS 800

Qy 116 TFIYIVPRVDAVDALKTCACLLKECRLEIAKCTANPSC-----AANVACIQTCNRPDET 169

Db 801 LAYLVEHAKATDH-----HLLSDSAVEELVKCKENPDMFEFLKEKSAKLGHTVWSNEAYSEL 856

Qy 170 ECQT-----KCGD--LFENSVVDFQFNECAVRKCKVPRKSDGVEFPVDPNRNAV 215

Db 857 EKKLEQPSLAYLVEHAKATDHLLSDSAVEDLVKC-----KENSDF-EF----- 899

Qy 216 VQNFNMKDFSGKWTTSGLNPTFDADFQCLHEFHMDKLVGNITWRIKTLTDGGFTTSA 275

Db 900 -----LREKSAKLGHTVWSNEAYSELEKLFQ-----PSLAYLVEHAKATDHLLSDSA 948

QY	276	VQTFVODPDLGALYNHONFEHLHQ-----DDWTILSQTENKPDPIFYVYKG	324
Db	949	YE-----DLVRCKENPDMEFLKESAKLGHTVVSNAYSELEKKLEQPSLEYLVSHAKA	1002
QY	325	RND--AWDGYGGSV-----IVTRSPTLPESTIIPN-----LQAKAASVGRDF---	363
Db	1003	TNHLLSDSAYEDLVKCKENPDMEFLKESAKLGHTVVSNAYSELEKKLEQPSLEYLVE	1062
QY	364	-----NNFITTONS-----CGPEPLVERLEKTAEBGEKLLIKEAV-EETEEVEKE---	408
Db	1063	HAKATNHLLSDSAYEELVKCKENPDVEFLKESAKLGHTVVSNAYSELEKKLEQPSLE	1122
QY	409	--VEKVRDTEMTLFORLLEGFKELQDEENFVRELSKEEKEILNE--LQMEA-TEVEKLF	463
Db	1123	YLVHEKATANHLLSD--SAYEELVKCKENPDVEFLKESAKLGHTVVSNAYSELEK--	1178
QY	464	GRALPIRKLRMALPHSNFLANHETIKYVGSKPLFGHKFRFSWGWEDYFGSIVVAK---	518
Db	1179	-----KLEQ--PSLAYLVEH-----AKATDHILLS--DSAYEDLVCKENPDV	1217
QY	519	-----ICSSRRIPRYFKPSRICCGLDSRGL-QLFSGHKHNLSPAHSINQNV	564
Db	1218	EFLUKESAKLGHTVVSNAYSELEKK-----LEQPSLAYLVEHAK--ATDHLLSDSA	1268
QY	565	PKGNSOCKFPKDVALMWKMGQFAKTAIVAFILSVASKAD-----AVDALKTCTCL	617
Db	1269	YEDLVCKENPDMEFLK-EKSAKLGHTVVSNAYSELEKKLEQPSLEYLVHEKATNHHL	1327
QY	618	LKECRLE-LAKCISNPAC-----AANVACLQTCNNRPDETECOIKCGDLFENSVDDEFN	670
Db	1328	LSDSAYEDLVKCKENPDMEFLKESAKLGHTVVSNKYESELEKKLEQPSL-----	1377
QY	671	ECAVSRKKCVPRK-SDVGDF-----PVPDPSVLVQKFDKDFSGKFWITRGLNPTDAF	723
Db	1378	EYLVKHAEQIOSKILISIDFNTLANPSMEDMASKLQLEYQIVSNDEVI--ALKNTMEKP	1435
QY	724	DCQ-----LHEFH-----TEENKLVGNLSWRITPDGFFTRSIVOKFTVQD--PKYPGLILYN	773
Db	1436	DVELLSRKLKGYHIDITTYNSELVSNFN-----SPTLKFEBAKSKGYRLI	1482
QY	774	HDNEYL-----LYQDDWYILSSK-----VENSPE-----D	798
Db	1483	EPNEYLDLNRIATPSPKEIDNFCQKICGYALDSKEYERLANSLENPKFEBENRAALD	1542
QY	799	YIFV---YKGRNDWDWGGSVLYTRSAVLPESITPIELOTAQOKVGRDENTFIKTDNTC	855
Db	1543	LVLVDKTEYQAMKD-----NASNKKSLIPSTKAL-----DFTM-----	1576
QY	856	GPEPLVERLEKKVBSGERTIIKEVEETEEVEKVRDKEVTFLSKLFEKGKELQD---	911
Db	1577	PAPQLASAEKSSLOK-----RTUSDIENELKAL-----GYVAIRKEMLPN	1616
QY	912	-ENBFLRELSKEEM-----DVLDDGLKMEATEVEKLF-----RALPIR	948
Db	1617	LERPIVDNASKNDVLNLCFSKLSIVPLSTEYDNMRKEHTKILNILDPSIDFLKECKEY	1676
QY	949	KLMVAETH-----CFTSPCHDRIRFSSDDGIRGLGIRKTRINGFLKILPPTQOSAD	1001
Db	1677	QMLTISIKHDEYEEKQEAENPGYEFTLEKASALGY-----ELVSEVELDMQKIDSPD	1729
QY	1002	LRTTGRSSRPLSAFRSGFSKGFIDIVPLPSKNELKELTAPULLKLGVGLACAFILVPSA	1061
Db	1730	IDYMQEKAARN-----EWVLL--RNEEKE-----ALQKKTEYPSLFLIEKAA	1770
QY	1062	-----DAVDALKTCACLLKGCRIELAKCIANPACAAANVACLQTCNNRPDETECQTKG	1114
Db	1771	GMNKKILVDQIEYDET-----IRKC-----NHPTRMLEBSCH	1802
QY	1115	DLFENSVDDEFNECAVSRKKCVPRK-SDLGEPAPDPSPVLVQNVENISDFNGKXWITSGLN	1173
Db	1803	HL--NLVLLDQOETSTYRLEPLBNRVNVEDINTLSKNTIAIPNTIYQDLIGKYE-----N	1855
QY	1174	PTFDAEDCQLHEFHTBGDNKLVGNISWRIKTLDSGFFTRSAVQKFDQPNQPGVLY----	1229

Db	1856	PNFDYLKDSLNK-----MDYVAISRQDYELMVAKYERQOLDYLKIS	1896
Qy	1230	-----NHONEYLHYODDWYILSLSKIENKPEDYIPVYVYGRNDWDG	1270
Db	1897	SEKIDHIVPLSEYNILMVTNYRNPISLYLKKEKAVLNHHILKEDDYKNIL-----AVSE	1950
Qy	1271	YGGAVVYTRSSVLPNSIIPELEKAAKISGRDFTTIRDTGTCPPALVERIEKTVEEGE	1330
Db	1951	HPTVILHSEKASLLNKVLVDK-----DFATMSRSIE---KPTIDFLSTKALSMG-	1997
Qy	1331	RIIVKEY-----EEIEEEVEKEVEKVGRTMTPLFORLAEGRNEUKQDBENVRLEUSKEEM	1385
Db	1998	KILWNETHKRNEKLLSEPDSEFLTMKAKEOGLIITISEKEYSELRDQIDRPSLDVLKEKA	2057
Qy	1386	EFIDEIKMEASEVEKLFGKALP	1407
Db	2058	AIFDSIIVENIEYQQLVNTTSP	2079
RESULT	15		
S54174			
C	Species:	Plasmodium falciparum	Plasmodium falciparum
C	Date:	08-Jul-1995	#sequence_revision 21-Jul-1995
C	Accession:	S54174	#text_change 09-Jun-2000
R	Tosh, K.T.; Kilbey, B.J.K.		
A	Description:	submitted to the EMBL Data Library, January 1995	
A	Reference number:	S54174	Isolation and characterisation of the topoisomerase 1 gene from
A	Accession:	S54174	
A	Status:	preliminary	
A	Molecule type:	DNA	
A	Residues:	1-839	<TOS>
C	Cross-references:	EMBL:X83758; NID:g790481; PID:g790482	
C	Superfamily:	eukaryotic type I DNA topoisomerase	
C	Keywords:	DNA binding; DNA replication; isomerase	

```

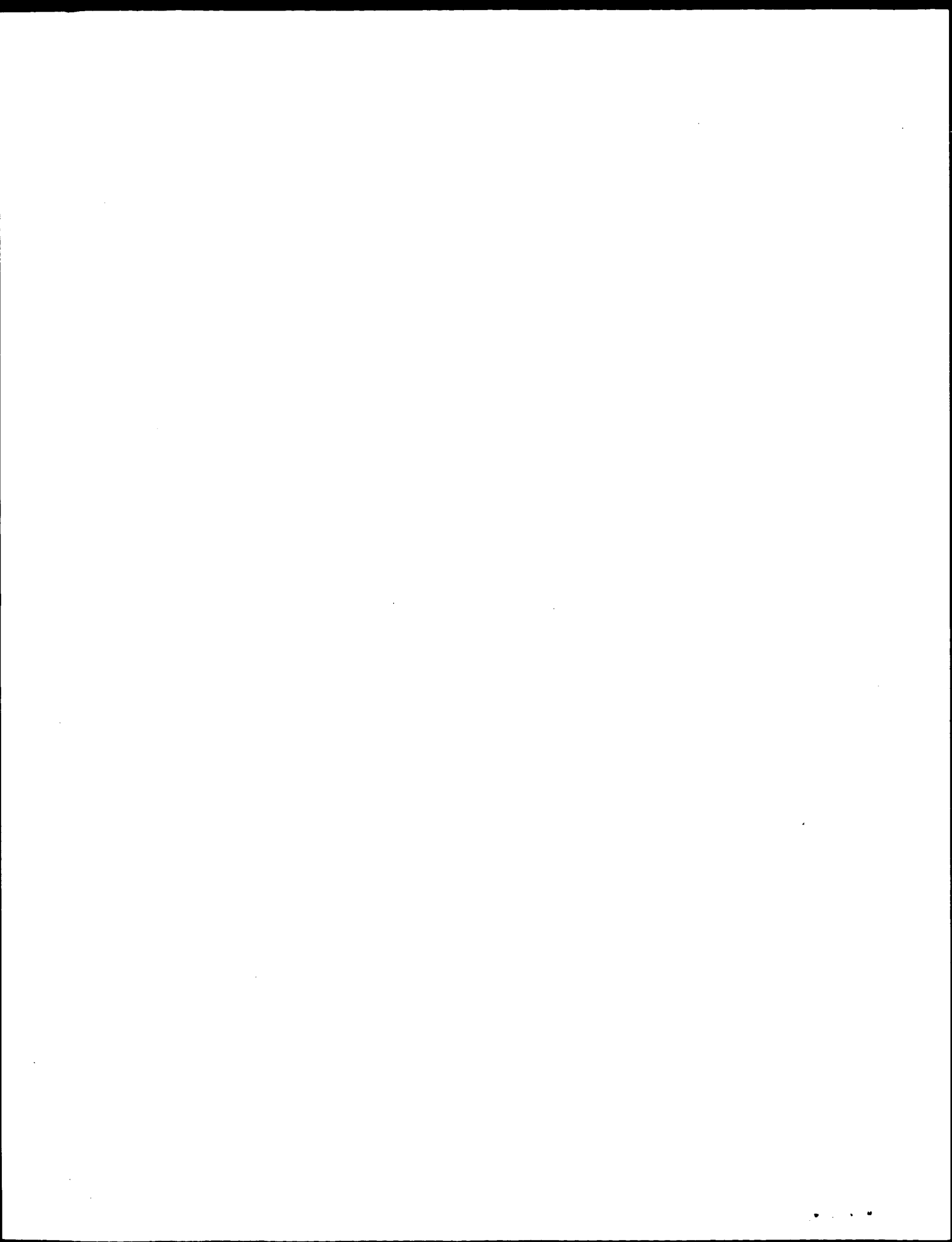
Query Match      1.9%; Score 144.5; DB 2; Length 839;
Best Local Similarity 17.3%; Pred. No. 1.1;
Matches 191; Conservative 176; Mismatches 390; Indels 347; Gaps
46;

QY 352 LOKAAKSVGRDFNFITFDNSGPEPPPLVERLEKTAEGEKLLIKEAVEI---EEVEKEV 409
      : : : : | | : : | | : : : : | | : : | | : : | | : :
Db 24 INK1QNLG---NN--KSCNRSRKSEIKKQKNSLGIKKNYKSLGIGKEEKKKQI 78
      : : : : | | : : | | : : : : | | : : | | : : | | : :
QY 410 EKVDRTEMTLFQRLLEGPELQODPEENVRELSKEEKEILNELQMEATEVFLGFRALPI 469
      : : : : | | : : | | : : : : | | : : | | : : | | : :
Db 79 SKRSNELKKNLKEGKKYVEKSRV---KDETCLTNVVKETQNNK-----PK 128
      : : : : | | : : | | : : : : | | : : | | : : | | : :
QY 470 RKLRLMALAPHNFANHETIIKYYGSKLPFGHKRISGWEDFGSVIVAKICSSRRIPRYF 529
      : : : : | | : : | | : : : : | | : : | | : : | | : :
Db 129 KILKKS---EENF-----EPINRW-WEKIDDDTDIQNW-----158
      : : : : | | : : | | : : : : | | : : | | : : | | : :
QY 530 RKSPPICGLDSRGLPESHGKHNLSPAHNS1QNVPKNSGCKPKPDVALMWEEKWGQFA 589
      : : : : | | : : | | : : : : | | : : | | : : | | : :
Db 159 -----LEHGL-IF-----SPPY-VQHVP-----IFKSIKIEL-----NA 188
      : : : : | | : : | | : : : : | | : : | | : : | | : :
QY 590 KTAIVAIFILSVASKADAVDAKTCUCLLKECLRELAKICSNAPCAANVACLOTNNRPD 649
      : : : : | | : : | | : : : : | | : : | | : : | | : :
Db 189 KSEELATWCS-----AIGSDYCTKEKFIILNFFKTFIN-----SLENDNIQK 231
      : : : : | | : : | | : : : : | | : : | | : : | | : :
QY 650 ETECOIKCDLFENSVDDEFNECAVSRKKVPRKSDVGDFVPDPDSVLVOKFDM-----703
      : : : : | | : : | | : : : : | | : : | | : : | | : :
Db 232 ENETKLKKGDISNFKFID-----FMP1KDHLLKRLREKLNKTEE 271
      : : : : | | : : | | : : : : | | : : | | : : | | : :
QY 704 KDFSCKWITIRGLNPTFDFAQCQLHEFHTEENKLVNLSWRIRTPDGG-----FFTRSAV 758
      : : : : | | : : | | : : : : | | : : | | : : | | : :
Db 272 KEEKKMRMERELPTYALVDWIREKISSNKAEPPLFRGRGHEPKQGLLKKRIFEDWV 331
      : : : : | | : : | | : : : : | | : : | | : : | | : :
QY 759 QKFVODPYPGILYN-----HDNE--YLLYQDDWYILSSKVENSPEDYIFVYK 806
      : : : : | | : : | | : : : : | | : : | | : : | | : :

```


Db 332 INISKDAPVRLYDNMCGHNGDIYHDNKVTLWAYKD-----SINDQIKYTFLS 381
Qy 807 RNDWDGYYGGSVLTSTRSAVLPEIPELOTAQKVGKGRDFNFIKT-----DNTCGPEPPLV 862
Db 382 AQSKFKGYKDLMKYENAR-----KLKSCVHKIREDYKNMKNNKIIDKQLGTAVYLI 433
Qy 863 ERLEKKVEEGERTTIKEVEEIEEVEKVRDEKVTFLSKFEGFKELORDEENFLRELSKE 922
Db 434 DFLALRV-GGEK-----DIDEADIV-----GCCSLRVEHISFAHDIPFK 472
Qy 923 EMDVLGLKMEATEVEKFLGRALPIRLKMAVATHCFTS---PCHDRIRFFSSDDGIGRLG 979
Db 473 SVD-----SKEQKTNDKVNKIPLTNLESISSEDCYITLDFLGKDSIRYENT-----520
Qy 980 ITRKRINGTFLKILPPIQSDADLTGTGGRSRPLSAPRSGFSKGIFDIVPLPSKNE-LKE 1038
Db 521 ---VKIDQAVYINII-----IFCKNKNRDEGVDFQITCSKLNEYLKE 559
Qy 1039 LTAPLLLLKLVGLACAFILVPSADAVDAL--KTCACLLKGCRIELAKCIANPACAANVAC 1096
Db 560 IMPILSAKVFTYNASITLDOQLKRIKEVYKTYTSLYSG-ETELHK-----SK 607
Qy 1097 LOTCNRRDETECOIKCGDLEFENSVDDEFNECAVSRK-----KCVPRKSDLGEPPADPSV 1152
Db 608 KRKSSHLTSDTNILSDASDSTINDVNNYDENGINKKLSYATTVGKENDVDKNSP-IEV 666
Qy 1153 LVQNFNISDFNGKWTYITSGLNPTDAFDCQLHEFHTGDNKLVGNISWRIKTLDSGFFTR 1212
Db 667 DVSNIN-----ELINFYNNANREVAAILCNHORSIPKQHDFTM 703
Qy 1213 SAVQKVVQDPNQPGVLYNHID-NEYLHYQDDWY ILSSKIENKPEDYIFVYVYGRNDAMDGY 1271
Db 704 SKIKKQIE-----LYNEDIKEYKY-----LQHLKKNSDKKFIFV-----738
Qy 1272 GGAVVYTRSSVLPNSIIPELEKAAKSIGRDFSTFIRTDNTCGPEPALVERIEKTVEGER 1331
Db 739 -----SKVSTLDGTLRP-----NKVKENMK 758
Qy 1332 IIVKEVEIEEVEKEVEKVGRTMTLQRLAEGFNLKQDEENFVRELSEMEFELDE- 1390
Db 759 -----EESCKKKL-----ITLIKVELLNQMKVRDDNKTTALGTSKINYMDPR 802
Qy 1391 -----IKMEASEVEKFLFGKALPIR 1409
Db 803 ITVAECKKFEIPIEKVFNRSRLK 826

Search completed: November 6, 2001, 05:02:57
Job time: 6721 sec




```
QY 337 IYTRSPLESIIPNIOAKASVGRDNFNFTTDSGCGPPPLVERLEKTAEGEKKLIK 396
Db 545 -----ISNHELSQSLTAAAEKAAVATNNE-----LSEKNSLQT 581
QY 397 EAVEIEEVEKEVEKVRDTEM-----TLFORLEGFKELOODEENFVRELSEKEEILN 450
Db 582 LCNAFEKIAKSVQKNEQNFSSLDTSFKKLNEHQLENNHQTITKQL-KDTSSKLO 640
QY 451 ELQMEATEVEK-----LFGRALPIRKLRLMALAPH-----SNFLANHETIKYVGS 495
Db 641 QLQERANFEQKESTLSDENNDLRTLLKLEESNKLKKEQDVDSLEKNIQTLK----- 695
QY 496 KLPCHKRFSGHWEDYFGSIYVAIKCSRRIPRYFRKSPRICGLDSRGLQLF---SHGKH 552
Db 696 -----ED-----LRKSEALRFSK-----LEAKNREVIDNKLKGH 726
QY 553 NLSPA-----HST-----NONVPKNGSGCKFKPKDVALMVWVKWGQPAKTAIIVAIFILSYA 602
Db 727 ETLEAQRNDLHSSLSDAKNINAILSELTKSSSDV-----KRLTANVETLQD 774
QY 603 SKA-----DAVDALKTCTCLLKECELEAKCISNPACAAVACLOTNNRPDETECOIK 656
Db 775 SKAMQSFSTLVNSYQISINLYHELRRDHVN-----MOSQNTLLESEKLIK 821
QY 657 --CGDLFEN--SVDFE-----NECAVSRKKCVPRK-----SDVGDFFVPDP 694
Db 822 TDENLTQONWTLIDNVQKLHKHVKVQESKSELKNGKSLDLKLNKSLNVAISDND 881
QY 695 SVLVQKEDM-KDFSGKWFTIRGLNPTDFDQOLHFEHTEE-----NKLGNLSWRIR 746
Db 882 QILTQLAELSKNYSLSQESQASLNGSLKLEAEKQLLHTNEELHRLDKLTGKLR----- 937
QY 747 TPDGFFTRSAVQKVFQDPKPYGILYNHNDXELLYLLODDMWILSKVENSPEYIFVYK 806
Db 938 -----IEESKSDL-----GKGLTARQEE-----ISNLKENMSQSAITSVKS 976
QY 807 RNDWDGYSGLVYTRSAVLESIIPELOTPAAQKVGRDNFTFKTONTGCGPEPLVERLE 866
Db 977 KLDE-----TLSSKSKL-EADIEHLKNKVEVEVERNALLASN-----ERLM 1017
QY 867 KKVE-EGERTIIEVEIEEVEKVRDKEVTLFSKL-----PEGF-----KELQD 911
Db 1018 DDLKNGE-----NIASLOTEIEKRAENDLQSLKSVSSEYENLLLISSQTNKSLD 1071
QY 912 BENFLRELSKEEMDVLGLKMEATEVEKL---FGR-----ALPIRKLMAVATHC 957
Db 1072 KTNQKLVYIEKNVQKLLDEKQDQNVLEELTSKYGKLGEENAAQIKDELLALRKK----- 1124
QY 958 FTSPCHDRIRFFSD-----DGIGRL-----GITRKRINGTFLKILPPIQS--AD 1001
Db 1125 -SKRQHDLCANFVDLKEKSDALEQLTNEKNELIVSLEQSNNEALVEPSDLNRLSD 1183
QY 1002 LRTTGGRRSPLSAFRSFGSKGIFDIIVPLPSKNELKELTAPLLKLVLGVLACAFVLPSA 1061
Db 1184 MKKLSDSNDVISIRSD-----LVRVN 1206
QY 1062 DAVDALKTACALLKGCRIELAKCIANPACAAVACLOTNNRPDETECOIKCGDLFENS 1121
Db 1207 DELDTLK-----KDKSLSTQYSEVCQDRDLDLSLKGK----- 1240
QY 1122 VDFNECAVS-RKKCVPRKSDLGEPAPDPVSVLVON--FNISDFNGKWYIT-SGLNPTFD 1177
Db 1241 EESFNKAVSLRELCTKSEIDV-----PVSEILDNDFVFNAGNFSELSRUTVLSLENYLD 1295
QY 1178 AFDQOLHEFHPEGNKLVGNISRIKTLDSGFTTRSAVQKVFQDPNPGVLYNHNDNEYLH 1237
Db 1296 AFN-QVNFKKMELDN-----RLTTDAEF-----TKVVDL-----EKLOH 1330
QY 1238 YDDWYILSSKIENKPEYIFVYTRGRNDWDGVGAVVYTRSVLPNSIIPLEAKKS 1297
Db 1331 EHDWLI-----ORG-----DLEKALD 1348
```

```
QY 1298 IGRDFTFIRDTCGPPEPALVERIEKTVEEGERILIVKEVEIEEVE----- 1345
Db 1349 SEKNF-----LRKEAWTENIH-SLEGKEETKKEIAELSSRLDNLQATNKLKNO 1398
QY 1346 -----KEY-----EKVGRTEMTLFORAEGFNELKODENFVRELSEKEMEFL--DEIKMEA 1395
Db 1399 LDHLNQETRLKEDVLKESLIISLEESLNQROKESLLD--AKNELEHMLDDTTSRKS 1456
QY 1396 SEVEKL 1401
Db 1457 SLMEKI 1462

RESULT 2
RBP1_PLAVB
ID RBP1_PLAVB STANDARD; PRT; 2869 AA.
AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN RBP1.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites.";
RL Cell 69:1213-1226(1992).
CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -!- SUBUNIT: HOMODIMER (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M8097; AAA29743.1; -.
DR HSP; P36956; IAW9.
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2869 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;

Query Match 2.1%; Score 155.5; DB 1; Length 2869;
Best Local Similarity 17.4%; Pred. No. 0.86;
Matches 283; Conservative 242; Mismatches 573; Indels 529; Gaps 76;

QY 57 TSYKTSFSDSSCHKDKSOICSDTSFEEI-----ORFDLKRGMTLILEKOWQFIOLA 110
Db 1150 TSNKNEILKSVKEVEDKLNVEQNEDEYKKNVKNPEKOLEAIRGSMKLKE----- 1200
QY 111 IVLVCTFVIVPVDVADALKTACALLKECRIELAKCIANPACAAVACLOTNNRPDTE 170
Db 1201 -----VINKHVSEMTQLESTANTLK-----SNAKGKENEHLEELN---KTK 1239
QY 171 COIKCGDLFE--NSVVDQFNECAVSRKKCVPRKSDVGEPVDRNA---VVQNFNM-KDF 224
Db 1240 GQWR--DIYEKLKKTAAELKEGTGVNELKDANEKANKVE-PEPERNIIGHVLERITVEKDK 1296
```

QY 225 SGKWTISLNPFTDAFCOLHEP--HMENKLVNLTWRIKTLDDGGFTTRSAVQTFVQD 282
Db 1297 AGK--VWEMNSLTKIEKLTQETSDSONELVTSTIKHLENAK----- 1340
QY 283 PDLPGALYNHDNEFLHYODDWYILSSQJENKPPDYIFVYVGRN-DAWDGYGGSVIVTRS 341
Db 1341 -----YED-----VTKNEEDSIQUREKAKSLETLD-----EM 1368
QY 342 PTLPSIIPNLOKAAK-----SVGRDPNN-----FITDMS----- 372
Db 1369 KKLVOQVNNLQSAIQGNAGISKELNELKGVIELLISTNYSSILEYVYKKNSSSVRESQL 1428
QY 373 CGPEPLVERLEKTAEE--EGEKLLKEAV-----ETEEVEK-----EVEKYRDT 415
Db 1429 ANGETTKAEGEKANASARLABAEKL--KEQIVKOLDYSDIDDKVKTEGIRKELRKMS 1486
QY 416 EMTLQRLLEGKELQOEDNEFVRELSKEEKEILN-----ELQME----- 455
Db 1487 ALTFWEE--SERFKQMCSSHMENAKEGKKIEYKNGDGKKANITDSQMEEVGNVVSKEA 1545
QY 456 -----ATEVEKLPGRALPIRKLMALPHSNFLANHETIKYY 492
Db 1546 HAFHTVEAQVDKTKAFCEISIVAYYTKMDNLFNLSL--MKEVKVCKE-----KNDEAEKY- 1598
QY 493 VGSKLPGHKRESWGWDYFGSIVVAKICSSRRIPRYFRKSPRICCGLDLSRGLQFSGKH 552
Db 1599 -SAKL-----KPYDGR- KARVSENERKISELKERAKV---EKKESSQLNDVSTK 1643
QY 553 NLSPAHSTNQVKNPGNSCKPKPDVALMWKVGQFATAIVAIFILSVASKADAVDALK 612
Db 1644 SILQIDNCROQLDSVLSNIGRVKQNALQYFDS-----ADKSMKSVLPISLGAESLDDVK 1699
QY 613 TCTCLLEKRELELAKCISNPACAAV--ACLOTNNRPDETECO-ICGDLFENSVDDEF 669
Db 1700 AA-----KEYEKNLETQONEMSRINVEGSLTDIDDKITDIENDLLKMKQYEEGLLOKI 1755
QY 670 NECAVSRRKKVPR-KSDVDPDPVPSVLVQKFDKDFSGKWFITRGLNFTDAFCOLH 728
Db 1756 KENADKRSNELVGESEINALLDPTSTIFI-KLKLKEYD-----MTGDLKNYGVKMN 1806
QY 729 EFHTEENKLVNLSWRI-----RTPDGGFFTRSAVQK 760
Db 1807 EIHGEFTKSNLYIETHLSNATDYSVTPEKAQSLRELAKEEHLRRREEREAIFLLNDIKK 1866
QY 761 FVQ-----DPKYPGLYNHDNEYLQYDWDWILS----- 789
Db 1867 VESLKLKEMMKVSAEYEGMKRDHTSVSQLVQDMKTIIVDELKTLNDISECSSVLNNVS 1926
QY 790 --SKVENSPE-DYIFVYVYKGRNDADWDYGGSVLYTRSAVLPESITPELQTAQKVGGRDPN 846
Db 1927 IVKVKESKHADY-----RRDANSWYESVITLANYFLSDEAKISS-----GMEFN 1971
QY 847 TFIKTD-----NYTCGPEPLVERLE-----KKVEGER-----TIKEY 880
Db 1972 AEMKSNFKTDLEBIFSVISNELLKIBQDSNDVIQERESEQLAKADATYVYNK 2031
QY 881 EEIEEEVEKVRDKEVTLFKLFEFGKELQDE-----ENFLRELSKEEMVDLGLKMEA 934
Db 2032 NEFNEKLEAAKNEEVSEKVRALKRLSOGVIRCHFENFRLDNT 2080
QY 935 TEVEKLFGRALPIRKLMVATHCTSPCHDIRFFSDDGIGRIGITRKRINGTFLKIL 994
Db 2081 -ELENL-----KKMTYIY-----RDKKSERESLQEM---ENEMN-TYNSI- 2117
QY 995 PPIQSADLRTGGRSSRPLSAFRSGFGKIDFIVPLSKNE-----LKELTAPILLKLVG 1050
Db 2118 --TOLEGIVVSAGESKE-----DIEKLSRNEEMRNISEKISTIDSK----- 2157
QY 1051 LACAFILVPSADAVALKTCACLLKGRIELAKCIANPACAAVACLOTNNRPDETECO 1110
Db 2158 -----VIEMNSTIDELYK---LGKNCQAHWISLISYATNMKTSKKLIMINKEKENTE-- 2206

QY 1111 IKCGDLFE--NSVVDEFNECAVSRKKCVPRKSDLG-EFFAPDPDSVLVQNFINISDFNGKWY 1167
Db 2207 -KVDYIKDNSSSTGDGYVETL-----KGFYGSKLTFSSASEIVQADTYSVNFAKH 2256
QY 1168 ITSGLNPFTDAFCOLHEPHEGDKNLVGNISWRIKTLDSGFTTRSAVQKFDVDPNPGV 1227
Db 2257 EKESLNAIRD- IKELYLFHONSISIV-----EGGVONML-----A 2292
QY 1228 LYNHDNEYLHYODDWYILSKIKENKPEDY-----IFVYVGRNDA----- 1267
Db 2293 LYDKLNEEKREMDLYRNISSETKLQWHEHSTOVFKPMIELHKGMMNETNKSLEKEKKLK 2352
QY 1268 -----WDYGGAVYTRSSV--LPN--SIIPELKAAKSIORD----- 1301
Db 2353 SVNDHMSMEAEIMKNGLKYTPESVQNNINNIYSVIEAEVKTLLEIDRDYGDNYQIVVEHK 2412
QY 1302 --FSTFRTDNTCGPEPALVERIEKTEVEGE-----RIIVKEVEEIEEVEKEVEKV--GR 1353
Db 2413 KQFSILLDRTN-----ALMDDEIEFKENNYNLMVNTETHRVNDYIEKTNKLVQAK 2466
QY 1354 TE-----MTLFOR---LARGFNEFKODEENFVRELSKEEM-----EFLDE 1390
Db 2467 TEYEILENIKQNDMLQNIPLFKKVSIIIEVFENVKSKESILNDLYEQERLLKIGHLD 2526
QY 1391 IKMEASE 1397
Db 2527 IKRNVT 2533
RESULT 3
ID TEGU_HSVSA STANDARD; PRT; 2469 AA.
AC Q01056;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE PROBABLE LARGE TEGUMENT PROTEIN.
GN 64 OR ERF2.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230228; PubMed=1314457;
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
organization between HVS and Epstein-Barr virus.";
RT Virology 188:296-310(1992).
CC -!- FUNCTION: TEGUMENT PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC EHVS-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64346; CAA45687.1; -;
DR EMBL; M86409; AAA46140.1; -;
DR PIR; H36812; H36812.

SQ SEQUENCE 2469 AA; 280165 MW; D2B4B8DC08644CDB CRC64;
 Query Match 2.1%; Score 155; DB 1; Length 2469;
 Best Local Similarity 17.3%; Pred. No. 0.75;
 Matches 230; Conservative 208; Mismatches 497; Indels 396; Gaps 58;
 QY 271 FTRSAVQTFVQDDPDLGALYNHDNEFLHVDQDWYILSSQIENKPPDYIFVYR----- 323
 Db 178 FTSNNDALIEYLSPPNVQYT--GSELYFVPEKEYIGHSH-----YIMNHRYVINYEKL 228
 QY 324 -GRNDWDGYYGVVY-----TRSPTLPESSIIIPNLQKAASVGRDFFNFI 367
 Db 229 HGNIDLTISOGELIIEISPPNTPKPTSTOKPKPTPTKPAKTPKPKTKPESTI 288
 QY 368 TTNSCGPEPLVERLEKAEAGEKLLIKAV-----EIEEVEKEVKVDT--EMTL 419
 Db 289 PYDKS--KKPP---KIPKTSKSKVLTDTALTPOKHITIEHLRELLPITETVEDNTL 343
 QY 420 FOR-----LLEGFKELQ-----QDENFVRELSKEKEILNELQMEATEVEKL 462
 Db 344 FNPVETPTGTSLSGINSITTKREDLEDNDNTSKLKEDEDDWIDD----- 392
 QY 463 FGRLPIRKLRLMALPHSNFLANHETIKYVVGSKLPGHKRFSGWEDYFGSIWA----- 517
 Db 393 ----IPIPEVLDTEHSS---DQETI-YMIGDE--NIHDWSYSDDDIDDTLIDISFIQLD 441
 QY 518 -KICSSRRIPRYRKSPRICGLDSRGLQFHSK-----HNLSPAHSNQNVKPG 567
 Db 442 NLITSLDNIPKN--NTEPRI---IDKTSNPIKREGKALHSIDRLKNIVLEHGLITSSSIS 497
 QY 568 NSGCKPKFDVALMWKQWGFQAKTAIVAFILSVASKADAVDAIKTCTCLLKECRELEAK 627
 Db 498 ISKCKSLQFVIL---WGE-----KLSIPTR-DLKTILKTELIITETAETALTK 542
 QY 628 CINSACAAVACLOTQNNRPDETECIQGGDLFENSVDNEFCAVSRKKCVPRKSDVG 587
 Db 543 -LTN-----DTRFNNTVITKLKMKMLK-----SESVD 569
 QY 688 DFP-----VPDPSVLQKFDN-----KDFS-----GKWFITRGLNPTF 720
 Db 570 SYRHLALLANNILKIOTIDTEIELKTLNVFTSELGKDFSVVCTKKESETIMAAIKNLK 629
 QY 721 DAFDQHLHEPHEENKLVGMLSWRIRPDGFFTRSAVOKF-----VQDPKPYG 769
 Db 630 EKISTRKQELTHEENYFQSVL-----IAMEFQPIPLTRVIEIQPSKKAQ 675
 QY 770 ILYNHDNEYLYODDWYILSSKVENSPEDYIFVYVYKGRND---AWDGYGGSVLYTRSAVL 826
 Db 676 QL--HEKSKLVEQ---LTIIDANNVLTLLHTMKQDKTDISAPD-----FTTVLKN 722
 QY 827 PESIPELOTAQKVGRD---FNTFTKTDNTCGPEPLVERLEKKEVEGERTI--IKEVEE 882
 Db 723 IQSTLQLLQTCVTDNLNKKFISNTVQOLSIVGWEVAELSHSNPPKADPVIPKLDD 782
 QY 883 IEEVEKVRDEKVTLSKLEFEGKELQORDENRELRSKEEMVDVLGKMEATEVEKLG 942
 Db 783 IKKEIQOVTTKQ-----KNEETLSKILADVOTLLENAKOSDTSILPILQHYITKATLVG 837
 QY 943 RALPIRLMAVATHCTSPCHDIRFSSDDGIGRLGIT----RKRINGTFLKLIPPIQ 998
 Db 838 E-----RENQFESLKNVTQKLSSTSEFLKTLIDSTLLENVQLIQ 878
 QY 999 S-ADLRTTGRSSRPLSAFRSGSKG---IFDIVPLPSKNEKELTAPILL----- 1045
 Db 879 EISDILOSNOYIHQSEFIKAFFDKSNTIINLTQINQOQKYTTVTQPMLIAVVKRFLSEA 938
 QY 1046 -----KLVGVLAACAFILVPSADAVDAIKTCAKLLKCGRIELAKCIANPACAAVAC 1096
 Db 939 KFRESNTICELISTVLSGLSKSTVTEALKDALKSIDLKEKL----- 983
 QY 1097 LQTCNNRPDETE-----CQIKGDLFENSVDNEFCAVSRKKCVPRKSDLGEPAPD 1149

Db 984 --TAVDRPLKRELYNVIRKLQKQLKTLLEQQEFDNW-----KMEVDSF-VPT 1027
 QY 1150 PSVLVQNF--NISDFNGKWIITSLNFTFAFCOLHEFHTEGDN-KLYGNISWRIKTLD 1206
 Db 1028 PSRDVKTFIONAPSKAKQYAKKALKQDIQAMEIDVDPEVIEDNLIKANGQKAWQ----- 1082
 QY 1207 SCFTFRSAVQKVFQDDPQNPQGVLYNHDNEXYLHYODDWYILSSKIENKPKEDYIF-----VY 1261
 Db 1083 -----KIQSAFODLNFSLI-----PDDWLSL-AKEVTRPKSTLFTVIGPILL 1124
 QY 1262 GRNDWDGYYGVVYTRSSVLPNSII-----PELEKAAK 1296
 Db 1125 KFVEEVLESVKNLKAEKLSLLPNGVPVTPPKFDWIHYHYESNVNFHLKTLNLPKVSVAH 1184
 QY 1297 SIGRDFSTFTRDNT-CGPEPALVERIE-----KTVE-----EGERIIVKEV 1337
 Db 1185 NIGHELSSLSOALNSKILPEAVGVTSLEQHAANFSCMFKLEATWHDHDOVDTTKIDEYI 1244
 QY 1338 BEIEEVEK-----EVEKVGK-----TEMTLFQRLAEGFNELKODEENFVRELKSEMEFL 1388
 Db 1245 EDLRNDTKKHIVAPQIQSPNRFSLPEDIQELNSLPKLFDRSLLENESRLASOKNEFQML 1304
 QY 1389 DEIKMEASEVE 1399
 Db 1305 -ENTVKAELQ 1314
 RESULT 4
 MYSB_PIG STANDARD; PRT; 1935 AA.
 AC P79293;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
 GN MYH7.
 OS Sus scrofa (pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 CX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DOMESTICA;
 RA Ko Y.L.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
 MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
 CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 DR EMBL; U75316; AAB37320.1; -
 DR InterPro; IPR000048; -

RT a comparative analysis of its product.";
 RL Genomics 8:194-206(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90301496; PubMed=2362820;
 RA Liew C.-C., Sole M.J., Yamauchi-Takahara K., Kellam B.,
 RA Anderson D.H., Lin L., Liew J.;
 RT "Complete sequence and organization of the human cardiac beta-myosin
 RT heavy chain gene.";
 RL Nucleic Acids Res. 18:3647-3651(1990).
 RN [3]
 RP SEQUENCE OF 1-115 FROM N.A.
 RX MEDLINE=89264452; PubMed=2726733;
 RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
 RT "Characterization of human cardiac myosin heavy chain genes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).
 RN [4]
 RP ERRATUM.
 RX Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;
 RA Proc. Natl. Acad. Sci. U.S.A. 86:7416-7417(1989).
 RN [5]
 RP SEQUENCE OF 1310-1935 FROM N.A.
 RX MEDLINE=86176778; PubMed=2421254;
 RA Saez L., Leinwand L.A.;
 RT "Characterization of diverse forms of myosin heavy chain expressed in
 RT adult human skeletal muscle.";
 RL Nucleic Acids Res. 14:2951-2969(1986).
 RN [6]
 RP REVISIONS.
 RX Leinwand L.A.;
 RA Submitted (MAR-1988) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 1410-1935 FROM N.A.
 RX MEDLINE=88299163; PubMed=2966919;
 RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;
 RT "Molecular cloning and characterization of human cardiac alpha- and
 RT beta-form myosin heavy chain complementary DNA clones. Regulation of
 RT expression during development and pressure overload in human
 RT atrium.";
 RL J. Clin. Invest. 82:524-531(1988).
 RN [8]
 RP SEQUENCE OF 785-1935 FROM N.A.
 RX TISSUE-Skeletal muscle;
 RC MEDLINE=90235862; PubMed=1691980;
 RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
 RA Arnold H.H.;
 RT "Identification of three developmentally controlled isoforms of human
 RT myosin heavy chains.";
 RL Eur. J. Biochem. 189:55-65(1990).
 RN [9]
 RP SEQUENCE OF 1393-1935 FROM N.A.
 RX MEDLINE=87192738; PubMed=3032769;
 RA Jandreski M.A., Liew C.-C.;
 RT "Construction of a human ventricular cDNA library and
 RT characterization of a beta myosin heavy chain cDNA clone.";
 RL Hum. Genet. 76:47-53(1987).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=96039076; PubMed=8533830;
 RA Arai S., Matsuo K., Hirayama K., Sukurai H., Tamura M., Ozawa T.,
 RA Kimura M., Imamura S.-I., Furutani Y., Joh-O K., Kawana M., Takao A.,
 RA Hosoda S., Momma K.;
 RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in
 RT hypertrophic cardiomyopathy.";
 RL Am. J. Med. Genet. 58:267-276(1995).
 RN [11]
 RP VARIANTS CMH1 GLU-256 AND ARG-741.
 RX MEDLINE=93248216; PubMed=8483915;
 RA Fananapazir L., Dalakas M.C., Cyran F., Cohn J., Epstein N.D.;
 RT "Missense mutations in the beta-myosin heavy-chain gene cause central
 RT core disease in hypertrophic cardiomyopathy.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).
 RN [12]
 RP VARIANTS CMH1 GLN-403.
 RX MEDLINE=90367131; PubMed=1975517;
 RA Geisterfer-Lowrance A.A.T., Kass S., Tanigawa G., Vosberg H.-P.,
 RA McKenna W., Seidman C.E., Seidman J.G.;
 RT "A molecular basis for familial hypertrophic cardiomyopathy: a beta
 RT cardiac myosin heavy chain gene missense mutation.";
 RL Cell 62:999-1006(1990).
 RN [13]
 RP VARIANTS CMH1 GLN-249; GLN-403; CYS-453 AND MET-606.
 RX MEDLINE=92204193; PubMed=1552912;
 RA Watkins H., Rosenzweig A., Hwang D.S., Levi T., McKenna W.,
 RA Seidman C.E., Seidman J.G.;
 RT "Characteristics and prognostic implications of myosin missense
 RT mutations in familial hypertrophic cardiomyopathy.";
 RL New Engl. J. Med. 326:1108-1114(1992).
 RN [14]
 RP VARIANTS CMH1 GLN-403; CYS-453; ARG-584 AND MET-606.
 RX MEDLINE=94070863; PubMed=8250038;
 RA Watkins H., Thierfelder L., Anan R., Jarcho J., Matsumori A.,
 RA McKenna W., Seidman J.G., Seidman C.E.;
 RT "Independent origin of identical beta cardiac myosin heavy-chain
 RT mutations in hypertrophic cardiomyopathy.";
 RL Am. J. Hum. Genet. 53:1180-1185(1993).
 RN [15]
 RP VARIANTS CMH1 GLN-403 AND VAL-908.
 RX MEDLINE=92346810; PubMed=1638703;
 RA Epstein N.D., Cohn G.M., Cyran F., Fananapazir L.;
 RT "Differences in clinical expression of hypertrophic cardiomyopathy
 RT associated with two distinct mutations in the beta-myosin heavy chain
 RT gene. A 908Leu->Val mutation and a 403Arg->Gln mutation.";
 RL Circulation 86:345-352(1992).
 RN [16]
 RP VARIANTS CMH1 LEU-403 AND TRP-403.
 RX MEDLINE=94075629; PubMed=8254035;
 RA Dausse E., Komajda M., Feltner L., Dubourg O., Dufour C., Carrier L.,
 RA Wisniewsky C., Bercovic J., Hengstenberg C., Al-Mahdawi S.;
 RT "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and
 RT identification of a hot spot for mutations in the beta-myosin heavy
 RT chain gene.";
 RL J. Clin. Invest. 92:2807-2813(1993).
 RN [17]
 RP VARIANTS CMH1 TRP-403.
 RX MEDLINE=94093568; PubMed=8268932;
 RA Moolman J.C., Brink P.A., Corfield V.A.;
 RT "Identification of a new missense mutation at Arg403, a CpG mutation
 RT hotspot, in exon 13 of the beta-myosin heavy chain gene in
 RT hypertrophic cardiomyopathy.";
 RL Hum. Mol. Genet. 2:1731-1732(1993).
 RN [18]
 RP VARIANTS CMH1 ASN-615.
 RX MEDLINE=93038688; PubMed=1417858;
 RA Nishi H., Kimura A., Harada H., Toshima H., Sasazuki T.;
 RT "Novel missense mutation in cardiac beta myosin heavy chain gene
 RT found in a Japanese patient with hypertrophic cardiomyopathy.";
 RL Biochem. Biophys. Res. Commun. 188:379-387(1992).
 RN [19]
 RP VARIANTS CMH1 GLY-778.
 RX MEDLINE=93343938; PubMed=8343162;
 RA Harada H., Kimura A., Nishi H., Sasazuki T., Toshima H.;
 RT "A missense mutation of cardiac beta-myosin heavy chain gene linked
 RT to familial hypertrophic cardiomyopathy in affected Japanese
 RT families.";
 RL Biochem. Biophys. Res. Commun. 194:791-798(1993).
 RN [20]
 RP VARIANTS CMH1 VAL-908.
 RX MEDLINE=93168485; PubMed=8435239;
 RA Al-Mahdawi S., Chamberlain S., Cleland J., Nihoyannopoulos P.,
 RA Gilligan D., French J., Choudhury L., Williamson R., Oakley C.;
 RT "Identification of a mutation in the beta cardiac myosin heavy chain
 RT gene in a family with hypertrophic cardiomyopathy.";
 RL Br. Heart J. 69:136-141(1993).
 RN [21]
 RP VARIANTS CMH1 TRP-719.

MEDLINE=95179132; PubMed=7874131;
Greve G., Bachinski L., Friedman D.L., Czernuzewicz G., Anan R.,
Towbin J., Seidman C.E., Roberts R.;
"Isolation of a de novo mutant myocardial beta MHC protein in a
pedigree with hypertrophic cardiomyopathy.";
Hum. Mol. Genet. 3:2073-2075(1994).
[22]
VARIANTS CMH1 CYS-513; ARG-716, AND TRP-719.
MEDLINE=94110336; PubMed=8282798;
Anan R., Greve G., Thierfelder L., Watkins H., McKenna W., Solomon S.,
Vecchio C., Shono H., Nakao S., Tanaka H., Mares A. Jr., Towbin J.A.,
Spirito P., Roberts R., Seidman J.G., Seidman C.E.;
"Prognostic implications of novel beta cardiac myosin heavy chain gene
mutations that cause familial hypertrophic cardiomyopathy.";
J. Clin. Invest. 93:280-285(1994).
[23]
VARIANT CMH1 THR-797
MEDLINE=96047159; PubMed=7581410;
Moolman J.C., Brink P.A., Corfield V.A.;
"Identification of a novel Ala797Thr mutation in exon 21 of the beta-
myosin heavy chain gene in hypertrophic cardiomyopathy.";
Hum. Mutat. 6:197-198(1995).
[24]
VARIANT CMH1 CYS-453.
MEDLINE=96209901; PubMed=8655135;
Ko Y.-L., Chen J.-J., Tang T.-K., Cheng J.-J., Lin S.-Y., Liou Y.-C.,
Kuan P., Wu C.-W., Lien W.-P., Liew C.-C.;
"Malignant familial hypertrophic cardiomyopathy in a family with a
453Arg-->Cys mutation in the beta-myosin heavy chain gene:
coexistence of sudden death and end-stage heart failure.";
Hum. Genet. 97:585-590(1996).
[25]
VARIANTS CMH1 THR-349 AND TRP-719.
MEDLINE=98204402; PubMed=9544842;
Jeschke B., Uhl K., Weist B., Schroder D., Meitinger T.,
Dontleumann C., Vosberg H.-P.;
"A high risk phenotype of hypertrophic cardiomyopathy associated with

DE DE NUCLEAR MIGRATION PROTEIN NUM1.
GN NUM1 OR YDR150W.
OS Saccharomyces cerevisiae (Baker's yeast).
GC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 28383 / FL100;
RC MEDLINE=92079907; PubMed=1745235;
RX Kormanec J., Schaaff-Gerstenschlaeger I., Zimmermann F.K.,
RA Pernecko D., Kuentzel H.;
RT "Nuclear migration in Saccharomyces cerevisiae is controlled by the
RT highly repetitive 313 kDa NUM1 protein.";
RL Mol. Gen. Genet. 230:277-287(1991).
CC -!- FUNCTION: CONTROLS NUCLEAR MIGRATION. NUM1 SPECIFICALLY CONTROLS
CC THE INTERACTION OF THE BUD NECK CYTOSKELETON WITH THE PRE-
CC DIVISIONAL G2 NUCLEUS PERHAPS BY RECOGNIZING G2-SPECIFIC
CC CYTOPLASMIC MICROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR
CC ENVELOPE.
CC -!- MISCELLANEOUS: ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT
CC CONSENSUS (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH
CC FLANKING DOMAINS OF THE TANDEM REPEATS.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61236; CAA43554.1; -.
DR PIR; S19052; S19052.
DR SGD; S0002557; NUM1.
DR InterPro; IPR001849; -.
DR Pfam; PF00169; PH; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Repeat.
FT DOMAIN 593 1384 12.5 X TANDEM REPEATS.
FT REPEAT 593 656 1.
FT REPEAT 657 727 2.
FT REPEAT 728 798 3.
FT REPEAT 799 862 4.
FT REPEAT 863 926 5.
FT REPEAT 927 990 6.
FT REPEAT 991 1054 7.
FT REPEAT 1055 1118 8.
FT REPEAT 1119 1182 9.
FT REPEAT 1183 1246 10.
FT REPEAT 1247 1310 11.
FT REPEAT 1311 1374 12.
FT REPEAT 1375 1384 13 (INCOMPLETE).
FT DOMAIN 2573 2683 PH.
SQ SEQUENCE 2748 AA; 313202 MW; B2FBD67C9F6211AE CRC64;

Db	857	EKKLEOPSLAYLVEHAKATDHHLLSDSAYEDLVKC-----KENS DV-EF-----	899
QY	216	VQNFNMKDFSGKWIYITSGNLPTTDAFCOQLHEFHENDKLVGNITWRIKTLDGGFTT RSA	275
Db	900	-----LKEKSAKLGHTVVSNEAYSELEKKLEQ-----PSLAYLVEHAKATDHHLLSDSA	948
QY	276	VQTFVQDPPDLPALYNHDNEFLHIY-----DDWYILSSQIOENKPDYIFVYYRG	324
Db	949	YE-----DLVKCKENPDMEFLKEKSAKLGHTVVSNEAYSELEKKLEQPSLEYLVVHAKA	1002
QY	325	RND-----AMDGYGGSV-----IYTRSPITLPESTIPN-----LOKAAKSVGRDF--	363
Db	1003	TNHLLSDSAYEDLVKCKENPDMEFLKEKSAKLGHTVVSNEAYSELEKKLEQPSLEYLVE	1062
QY	364	-----NNFTITNS-----CGPEPPIVERLEKTAEBEGEKLLIKEAV-EIEEVEVEKE--	408
Db	1063	HAKATNHLLSDSAYEELVKCKENPDVEFLKEKSAKLGHTVVSNEAYSELEKKLEQPSLE	1122
QY	409	--VEKVRDTEMTLFORLLEGKELQODEENFVRELSBEEKELINE--LOMEA-TEVEKLF	463
Db	1123	YLVEHAKATNHLLSD--SAYEELVKCKENPDVEFLKEKSAKLGHTVVSNEAYSELEK--	1178
QY	464	GRALPTRKURMALAPHSNPLANHETIKYVGSKLPGRKFRSGWEDYFGSIVAVK-----	518
Db	1179	-----KLEQ-----PSLAYLVH-----AKATDHHLLS-----DSAYEDLVKCKENPDV	1217
QY	519	-----IGSSRRIPRYFKSPRICGLDSRGL-OLFSHGRHNLSPAHISNQNV	564
Db	1218	EFLKEKSAKLGHTVVSNEAYSELEKK-----LEQPSLAYLVEHAK--ATDHHLLSDSA	1268
QY	565	PKONGCKTFPKDVALMVWEKQGFAKTAIVAFILSVASKAD-----AVDALKTCTCL	617
Db	1269	YEDLVKCKENPDMEFLK-EKSAKLGHTVVSNEAYSELEKKLEQPSLEYLVHAKATNHLL	1327
QY	618	LKECRLE-LAKCISNPAC-----AANVACLQTCNNRPDETECOIKCGDLFENSVVDEFN	670
Db	1328	LDSAYEDLVKCKENPDMEFLKEKSAKLGHTVVSNEAYSELEKKLEQPSL-----	1377
QY	671	ECAVSRKKCVPRK-SDVGDF-----PVPDP-SVLVQKDFDKSGKWFITRGLMPTDFAF	723
Db	1378	EYLVKAEQIOQSILISDSFNTLANPSMEDWASKLQKLEYQIVSNDEYI--ALKNTMEKP	1435
QY	724	DCQ-----LHEPH-----TEENKLVGNLSRITPDGFGFTTSVAKQFVOD-PKYPGILYN	773
Db	1436	DVLLRSKLKGVIHIDTTYNELVSFN-----SPTLKFEKAKSGKFRLI	1482
QY	774	HDNEYL-----LYQDDWYILSSK-----VENSPE-----D	798
Db	1483	EPNEYLDLNRIATTPSKEIDNFCQIGYALDSKEYERLANSPNSKFFIEBENAALD	1542
QY	799	YIFV---YKGRNDWDGVGGSVLYTRSAVLPESILPELOATAQKVGDFNFTFKTDNTC	855
Db	1543	LVLVDKTEYQAMKD-----NASNKSLIPSTKVL-----DFVTM-----	1576
QY	856	GPEPPIVERLEKKVBERGTIIKEVEETEEVEKVRDKEVTILSKLPEGKELQORD---	911
Db	1577	PAPQIASAEKSSLQ-----RTLSDIENELKAL-----GYVAIRKENLPN	1616
QY	912	-BENFLRELISKEEM-----DVLGLKMEATEVEKLFG-----RALPIR	948
Db	1617	LERPIVDNASKNDVLNCSKFSVLPLSTEEYDNMRKEHTKILNILDGSPIDFLKEKCEKY	1676
QY	949	KMAVATH-----CFTSPCHDRIRFTSSDDGIGRLGITRKRINGITFLKILPPIOSAD	1001
Db	1677	QMLIIKSHDYEEKQAEIENPGYEFILEKASALGY-----ELVSEVELDRMKQOMIDSPD	1729
QY	1002	LRTTGGRSRPLSAFRSGFSKGFIDFVLPLPSKNELKELTAPLLKLVLGVLCAPLIVPSA	1061
Db	1730	IDYWOEKAARN-----EMVLL--RNEEKE-----ALQKKIEYPSLTELIEKAA	1770
QY	1062	-----DAVDALKTACILKGCRIELAKCTIANPACAAVACLQTCNNRPDETCCQIKCG	1114
Db	1771	GMNKILVDDQIETDET-----IRKC-----NHPTRMSELSCH	1802

QY 506 GWEDYFGSIVVAKICSSRRIPRYKSPRCCGLDSRGQLQFSHGKHNLSPAHSINQNPV 565
Db 402 HLQNCQIETFTVELEHKVPIINSEKERTMDLENELNNAALLLHTSNEKNAKVKELN--A 458
QY 566 KGNQCGKFPKDVAMVWEKW-----GQFAKTAIVAIFLTVASKA 605
Db 459 KNOKLVECDLQTLTKORLDLCRQIYLLITNSVNSDKGPKLKEIQ--FIONIQMED 516
QY 606 DAV-----DALKTCCLLKECR--LEIAKICSNPACANVACIQTCCNNRPDETCQKCG- 658
Db 517 DSTITESDSQKVTVERLVEFKNIQLOE-----KNAELLKVRNLRADKLESKEKSK 568
QY 659 --DLFNSVVDENECASVKCKVPEKSDVGD-----FPVPDPSPVLVQKFD 702
Db 569 QSLQKIESEITVNEAKEAILTKS---EKMDSRIEELQELKEELTKSVPNEDASVNT 625
QY 703 MKDFSGKWFITRGLNPFTFADFQOLQHEFHTENKGLNLSWRITPDGGFFTRSAVQKVF 762
Db 626 IKQ-----LTETKRDLQSVQDLQTRISOIT-----RESTEN-----MSLLNKEI 665
QY 763 QDPKYPGILYNHNEY-----LLYQDDWYILS-----SKVNSPEDYIFVYKYG 806
Db 666 QD-----LYDSKSDISIKLGKSSRIAEERFKLLSNTLDLTKAENDQLRRFPDYLQN 719
QY 807 -----RNDWDGYYG-----SVLYTRSAPVLPESIIPELQTAQKVGGRDFNFIKTD- 852
Db 720 TILQDQSKTHETLNEYYSCKSKSIVETELLNKEP-----OKLRVHLEKNLQKEL 770
QY 853 NTCQPE-----PPIVERLEKKVEGERTIIEVEIEEVEEKVKRDEKVTFLSKL 901
Db 771 NKLSPEDSLRIMVTLQTLQKEREDELEETKRSCKQKIDEDALSSELK-KET-----SOK 826
QY 902 FEGKELQORDE-----ENFLRELKEEMDVLGDKMEATEVEKLFGRALPRKILMAVA 954
Db 827 DHHKQLEEDNNSNIEWQNKIEALKADYESVTSVDSKQTDIEKQYKVSKEKEIE-- 884
QY 955 THCFPSCHDIRFPS-----SDGIGRLGITRKRINGTFLKLPILPQIASDLRTTG 1006
Db 885 -----EDKIRLHTYVMDETINDDSL-RKELEKSKINLTDAYSQIK--EYKDLVET- 932
QY 1007 GRSSRPLSAFSGFSKGFIDIVPLPFSKNELKELTAPLLLVGLVACAFIVPSADAVDA 1066
Db 933 --TSQSLQQTNSKLUDESKFOFT-----NQLKNTL-----DEKTS 964
QY 1067 LKTCACLLKGRICRIELACIANPACANVACIQTCCNNRPDETCQIKGDLFEN-----SV 1121
Db 965 LEDKISLLKEQMFNL-----NNELDLQKGMKEKADFKKRISILQNNNKEVAV 1014
QY 1122 VDFNECAVSRKKCVPRKSDLGPEPAPDPSPVLVQNFNISDFNGKWIYTSGLNFTTFAFDC 1181
Db 1015 KSEY-----ESKLSKTQNDL-----DOQTIYANTQANNYEQELQKHADVSKTISELRE 1062
QY 1182 QLHEFHTEGDKNLVGNISWRIKTLDGSGFFTRSAVQKVPQPNQPGVLYNHDNLYHYQDD 1241
Db 1063 QLHY-----KGQVKTLN-----LSRDQLNALKNEKS-----WSSQKESLLEQLD 1104
QY 1242 WYLUSSKIENKPE-----DYIFVYVYGRNDAMDYGG-----AVYVTRSSVLNPSIIP 1289
Db 1105 --LNSRIEDLSSQKLLYDQIQIYTAADKEVNNSTNGPGNLLITLRRERDILDTKVT 1162
QY 1290 ELEKAASIGRDFSTF-----IRTDNCGPEPALVERIEKTVETGERIIVKEVEEIE 1341
Db 1163 VAERDAKMLROKISMDVQLQDARTKLDNS-----RVKE-----ENHSIIQQHDDIM 1210
QY 1342 BEVEKEVKGRTMTLQRLAEGFNELKODEENFVRELSEKEEMFEIDKMEASEVE 1399
Db 1211 EKL-NQLNLLRESNITL-----RNEL-ENNNKKKELQSE-----LDKLNQNVAPTE 1255

RESULT 10

DMD_CHICK

ID_DMD_CHICK

STANDARD;

PRT; 3660 AA.

Query Match

Best Local Similarity

1.8%; Score 135; DB 1; Length 3660;

Pred. No. 17;

AC DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DYSTROPHIN.
GN DMD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCHI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098331; PubMed=3062582;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "Nucleotide sequence of chicken dystrophin cDNA";
RL Nucleic Acids Res. 16:11815-11815(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Muscle;
MEDLINE=89210800; PubMed=3072195;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "The chicken dystrophin cDNA: striking conservation of the C-terminal coding and 3' untranslated regions between man and chicken.";
RL EMBO J. 7:4157-4162(1988).
CC -!- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE PLASMA MEMBRANE.
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN, ABP-120, ABP-180, OR BETA-FODRIN).
CC -!- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL; X13369; CAA31746.1; -
DR PIR; S02041; S02041.
DR HSP; Q01082; 1A2.
DR InterPro; IPR000433; -
DR InterPro; IPR001202; -
DR InterPro; IPR001589; -
DR InterPro; IPR001715; -
DR InterPro; IPR002017; -
DR InterPro; IPR002349; -
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00397; WW; 1.
DR Pfam; PF00569; Z2; 1.
DR Pfam; PF00435; spectrin; 22.
DR PRINTS; PR00403; WWDOMAIN.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; 1.
DR PROSITE; PS00021; CH; 2.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
KW Repeat.
FT DOMAIN 1 244 ACTIN-BINDING.
FT DOMAIN 300 3000 26 SPECTRIN-LIKE REPEATS.
FT DOMAIN 3052 3085 WW.
FT DOMAIN 3086 3357 CYS-RICH.
FT VARIANT 1171 1171 MISSING.
FT VARIANT 1869 1869 Q -> H.
FT VARIANT 1885 1885 K -> R.
SQ SEQUENCE 3660 AA; 422874 MW; 85493DAF6D5B6D4A CRC64;

Matches 245; Conservative 196; Mismatches 482; Indels 472; Gaps 64;

```
QY 71 KDKQICISDTSFBEIQRFDLKRM--TLI-LEKQWRF-----IQLAIVLC 115
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1923 KKKEDLNAVNRQAERLSKDGAAKAVEPTLVQLSKWRDFESKFAQFRRLNYAQITVLED 1982
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 TFIIVPRVDAVALKTCACLLKCEIRIACIANPSCAANVACLOTCNNRPDETEQIKC 175
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1983 TTFVWTSMTVETTYVSTYLAET-IQLQLALSEVERLNSPVLQAKD-----C 2030
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 GDLFENSVDQFNCAVSRKKCVPR-----KSDVGEFPPVDRNAVQFNKNKDF 224
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2031 EDLLKQ-----ECLKNIKDCLGRLOGHDIHISKTPALQSATPRETANIOD-KLTQL 2083
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 225 SGKWIYISGL-----AFDCQLHEFH-----MENDKL-----VGNLT 260
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2084 NSOWEKYNNKYRDRQARFDRSKSEKWRFLHCEMKSFNELWTEBEKLSRAQIEAGDVGHVK 2143
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 261 WR--IKTLDGGFTFRSAVQTFVQDDPLPG-----287
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2144 TKQFLQELQDGI---GRQQTVVKTNLNTGTEGEEIIEQSSAADANVLKQGLNLTNRWQEI 2200
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 288 -----ALYNHNDNEFLHYODDYILSSQIENKPDYIFVYGRNDAMDGYGSGVIYTR 340
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2201 QLVKRRRIEENKILSEFQEDLNKLLIWEETENVTAIPLPGNEDQLPDCGLGKVKLRV 2260
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 341 SPTLPESIIINLOKAASVGRDNFTTNSCGPPEPLVERLEKTAEBGEKLLIKEAVE 400
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2261 EELLPHKGI--LKLNETG-----TTLSASLANPERKHLESTLKASRLLKVSRD 2311
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 401 IEVEKEVEKVRDTEMTLQRLLEGFKELQDDEENFVRELSKEEKEIINELQMEATEVE 460
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2312 LPEK-QKEIE-----ILLKDFIELNQ-----INQLTWITPVK 2344
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 461 ---KLGR-----ALPIRKLRMALAPHNSFLANHETIKYVGSKLPGHKRFSWGWDYFG 512
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2345 NQLELYNQVGOPGAFDIKTEAA-----VQAKQPNVE-----2377
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 513 SIIVAKICSSRRIPRFRKSPRICCGLDGSLGQLFSGHKHNLSPAHSINQNVPKGNSGCK 572
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2378 --VLKSGC-----HLYKEKP-----ATHPVKKKLEDNLADWK 2407
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 573 FPKDVALMWEKMGQAKTAIVAIFILSVASKADAVDALTKTCTCLKECRLELAKCISNP 632
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2408 AINHLIILQLEK-PTGEPALTPSGVLT-SGQIVADT-----QARVYKETSFTTP 2456
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 633 ACAANVACLOTCNNRPDETEQIKCGDLFENSVDENEC-----AVSR--KKCVPRKS 684
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2457 EMPSSV-----LLEVPALADFNKAWAELTDWLSKLDREITKAQRV 2495
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 685 DVGDFPVPDPSVLVOKFDMKDFSGKWFITRGLNPTTDAFDCQLHEPHEEENKLVGNLSWR 744
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2496 TVGDLDDINDMIITQKANMODLEQR-----RP-----OLDELITAAQ-----NLKKN 2537
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 745 IRTPDGFGFTFRSAVQKF-VQDPKYPGILYNHNE-YLLYODDWYILSSKVENSPEDYIFV 802
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2538 TSNOEARTIITDRIEKIQSQWDDVHGLQNRQOLHEMOKDSTOWLEAKQE---AEQVLE 2594
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 803 YYKGRNDAMDGYGSGVLYTRSAVLPSI-----IPELOTAQAKVGRDFN- 846
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2595 QAKAKLESWK-----EISYTVALKKQNSSELKQFSKEIROMQNMIEGVNDVALKPVVDYSA 2650
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 847 -----TFIKTDNTCGPEPLVERLEKKVEEGERTI-----IKEVE 881
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2651 DDTKVELMTDNI-----NATWATINKRVSEEALESALLMLQEFYLDLEKFLAMLTEAE 2706
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 882 EIEEVEKVRDTEVTL-----FSKLEGEKELQD-----EEN---FLRELSKE 922
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2707 TTANVLQDATHKEKTLEDQPMVRELKMQQDLQAEIDAHTDIFHNLDENGQKILSLEGS 2766
| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 923 EMDV-----LDGLKMEATEVEKFLGFRALPIRKLMVATHCFTSPCHDIRTRFSSDDIGIR 977
| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2767 EDVAVLLQRLDNNMFRWSELK---KSLNTRSHLEAT-----DQWKR 2806
| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 11

MYSP_HUMAN

```
ID MYSP_HUMAN STANDARD; PRT: 1937 AA.
AC P13535; Q14910;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MYOSIN HEAVY CHAIN, PERINATAL SKELETAL MUSCLE.
GN MYH8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=90323631; PubMed=2373371;
RA Karsch-Mizrachi I., Feghali R., Shows T.B. Jr., Leinwand L.A.;
RT "Generation of a full-length human perinatal myosin
heavy-chain-encoding cDNA.";
RL Gene 89:289-294(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=95324556; PubMed=7601129;
RA Jullian E.H., Kelly A.M., Pompidou A.J., Hoffman R., Schiaffino S.,
Stedman H.H., Rubinstein N.A.;
RT "Characterization of a human perinatal myosin heavy-chain
transcript.";
RL Eur. J. Biochem. 230:1001-1006(1995).
RN [3]
RP SEQUENCE OF 502-1937 FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=90235862; PubMed=1691980;
RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
Arnold H.H.;
RT "Identification of three developmentally controlled isoforms of human
myosin heavy chains.";
RL Eur. J. Biochem. 189:55-65(1990).
RN [4]
RP SEQUENCE OF 860-1937 FROM N.A.
RX MEDLINE=89234168; PubMed=2715179;
RA Feghali R., Leinwand L.A.;
RT "Molecular genetic characterization of a developmentally regulated
human perinatal myosin heavy chain.";
RL J. Cell Biol. 108:1791-1797(1989).
RN [5]
```


Db 1113 ELQARIBELGEIEAERASRAKAEQKQSDLSRELEISERLEEAGGATSQAQVELNKKREA 1172
QY 890 -----VPDKVTLFSLKFEQKELQDEENFLRELSKEEMDVL 927
Db 1173 EFQKLRDLLEATLQHAMVAALRKHADSMALGEIDNLQVRKOKLEKSELKMET- 1231
QY 928 DGLKMEATEVEKLFGRALPIRKILMAVATHCFTSPCHDIRFFSSDDGIGRLGITRKRING 987
Db 1232 DDLSSNAEATSKAGN--LEKM-----CR-----SLEQVSEL--KTREEQ 1269
QY 988 TFLKILPPTQSDLRRTGGRSRPL---SAFRSGFSKGFIDIVPLPSKNEKELTAPLL 1044
Db 1270 QRLINDL-TAQRALQTEAGEYSRQLDEKDALYSQLSRS-----KQASTQOIEELKHOLE 1323
QY 1045 LKLVGLACAFILVPSADAVALKTCACLLKGCRIELAKCIANPACAAVACLOTNNRP 1104
Db 1324 EETAKNALAHALOSSRHDCDLLEQYEEQEGKAEQLQALSK--ANSEVAQWRT----- 1376
QY 1105 DETECQIKCGDLFENSVVDEFNECAVSRKCVPRKSDGLGEPFAPDPFVLQVNFNISDFNG 1164
Db 1377 -----KYETDAIQTEELEEAKKLAQRLQEAEE-----HVEAVNA 1412
QY 1165 KWTITSLGNTFFDAFCQLHEFHTEGDNKLVGNISMRIKTLDSGFFTSRAVQKRVQDPNQ 1224
Db 1413 K--CASLEKTQKQLQNEVEDMLDVE-----RSNAACAALDKKQ 1449
QY 1225 PGVLYNHDNEYLHYQDDWYILSKKIEKNKPEDYIFVYGRNDAMDWDGYGAVVYTRSVLP 1284
Db 1450 ----RNFDKVLSEMKQYEEQAELEASQKE-----SRSLTELFKVKNYVEESL--- 1495
QY 1285 NSIPELEKAKSIGRDFSFRTDNTCGPEPALVERIEKTVEEGRIVKEVEETEEV 1344
Db 1496 -DQLETLRNKNLQOEISDLTQIAEGGQIHELEKIKQVQEQKEQIAQALEEAEASL 1554
QY 1345 EKEVEKVGRTMTLFO-----RLAEGFNLKQDEENFVRELS----- 1381
Db 1555 EHEGKILRIQLELNQVSKSEVDRKIAEKDEIDQLKRNHTRVVTMQSTLDABEIRSRNDA 1614
QY 1382 ---KEEME-FLDEIKMEASEVEKLFCKAL 1406
Db 1615 LRVKMKMGDLNEMEIQLNHNALAEASL 1643

RESULT 12
MSPL_PLAFK
ID MSPL_PLAFK STANDARD; PRT: 1630 AA.
AC P04932;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMSEA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_Taxid=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RA Pan W., Tolle R., Bujard H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POTENTIAL).
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC -----This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03371; CAA27070.1; -
DR PIR; A25120; SAZQK1.
DR InterPro; IPR000561; -
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 1630
FT DOMAIN 67 84
FT TRANSMEM 1614 1630
FT CARBOHYD 97 97
FT CARBOHYD 259 259
FT CARBOHYD 755 755
FT CARBOHYD 759 759
FT CARBOHYD 774 774
FT CARBOHYD 835 835
FT CARBOHYD 911 911
FT CARBOHYD 955 955
FT CARBOHYD 1049 1049
FT CARBOHYD 1156 1156
FT CARBOHYD 1165 1165
FT CARBOHYD 1436 1436
FT CARBOHYD 1517 1517
SQ SEQUENCE 1630 AA; 187289 MW; ADBEC3CE0A46322 CRC64;
Query Match 1.8%; Score 134; DB 1; Length 1630;
Best Local Similarity 18.0%; Pred. No. 6.7;
Matches 246; Conservative 207; Mismatches 473; Indels 444; Gaps 70;
QY 184 VDOFNECAVSRKKCVPRKSDVGEFFVDPDRNAVQNFNMKDFSGKWIYTSGLNPTFDAFC 243
Db 240 IENINELTEESKKTIDK-----NKNATKEEKKLYQAQYDLS-----IYNK 281
QY 244 QLHEFHMDNDKLVGNLTWRIKTLDCGFFTRSAVQTF--VQDP-----DLPCALYNHNE 295
Db 282 QLEEAH----NLISVLEKRIIDTLKKNENIKELLDKINPPPPANSNGTNTLLDNKK 337
QY 296 FLHYQDDWYILSSQIENKPDY-----IFVYGRNDWDGYSVYVTRSPILPESII 349
Db 338 IEHEKEKEIKETAKTKFNIDSLFTDPLEYLYLRKKNID--ISAKVETKESTEPN-Y 394
QY 350 PNLQAKAASVGRDFNNFTTDSNCGPEPLVERLEKTAEEGEKLLI-----KEAVEIEEE 404
Db 395 PNGVYPLSY-NDINNALNELNSFG---DLINPFDTKEPSKNIYTDNERKKFNEIKEK 450
QY 405 VEKEVEKVRDTEMTLFQRLIGCFKELQDEENFVRELSKEEKEILNEL-----QMEAT 457
Db 451 IKIEKKKI-----ESDKSYEDRSKSLNDITKEYEKLLEIYDSKFNNDILT 498
QY 458 EVEKEKLFGR--ALPIRKRLMALPHSNFLANHETIKYVYVGSKLPGHFRFSWGWEDY-FGSI 514
Db 499 NFEKMGKRYSYKVEKL-----THHTFASYENSKHNL-EKLTALKY--MEDYSLRNI 549
QY 515 VVAKTCSRRRPRYPRKSPRCCGLDSRGLQFSGHGNLSPAHISINONVPKSGSGC--- 571
Db 550 VVEKEL-----KYYK-----NLISKIENEI---ETLVENIKKDEQLFEK 586
QY 572 -----KPKPDVALMVWEKQGFAKTAIVAFILSVASKADAVALKTCCTLLCEKLE- 624
Db 587 KITKDNKPDKEILEV-----SDIVKVVQVKVLLM-----NKIDELKTKQLILKRNVELKH 636
QY 625 -----LAKISN-----PACAA-----NVACLOTGNNRPDET 651

Db 637 NIHVPNSYKQENKQEPYILVLKKEIDKLKVFMPKVESLINEEKKNIKTQEGSDNSEPST 696
QY 652 ECQI-----KGDILFENSVDNECAVSRKKVPRKSDVDGDFVPDPVLUQ-- 699
Db 697 EGEITQATTKPQQAGSALLEG---DSVQAQAEQKQAQP-----PVPVPPEAKAQP 747
QY 700 -----KFDMDKFSKWEITRGLNPTDFADQCLHEFTHEENKLVGLNSWRI 745
Db 748 TTPAPVNNKTENYSKLDYLE-----KLYEFL-----NTSYIC 779
QY 746 RTPDGGFFTRSAVOKFVQDPKYPGILYNDH--NEYLLYODDWYLLSKVEN-----SPED 798
Db 780 H-----KY--ILVSHSTMNEKILQ--YKITEESKLSLSCDPLD 815
QY 799 YIF-----VYVYGRNDWDGYSV-----LYTRSAVL-----PESIIPELQTAQ 839
Db 816 LFNINQNIIVM-----SMFDSLNLSQLFMEIYKEMVCNLYKLDKNDKKNLLEAK 871
QY 840 KVRGDFNTFIKTONTGPEP-PLVERLEKKVEGERT-----IKRVEEI----- 883
Db 872 KV-----STSVKTLSSSMQPLSLTPQDKPEVSANDDTSHSTNLNLSKLLENILSLGKKN 927
QY 884 -----EEVEKVRDKVETLFSKLFEGFKELORDEENFLRSELSKEEMDVLGK 931
Db 928 NIYQELIGQKSSNFYKILKDSOTFYNFVSKADDDINSLSDESKR-----K 980
QY 932 MEATEVEKFLGRALPIRKLMAVATHCTSPCHDRIRFFSDGIGRLGTRKRINFTLL 991
Db 981 LE-EDINKL-----KKTQLSFDLYNKKYKLERLFDKKTGVKYKMQIKLT---LL 1029
QY 992 KILPPIOSADLRTTGGRRSP---LSAFRSFGSKGIFDIVPLPSKNEKELTAPLLKLV 1048
Db 1030 K-----EQLESKLSNPNKPHVLFQNFVFNK-----KKEAE----- 1061
QY 1049 GVLACAFIIVPSADAVDAKTCALTKGCR--TELAKCIANPACAAVACLOTCNNRPDE 1106
Db 1062 -----IAETENTLTKLKHLYKLYVYNGESSPLKTLSESIQETEDNYASL 1110
QY 1107 TECOIKGDLFENSVDNECAVSRKKVPRKSDIAGEFPAPDPVSLVNFNDSFNGKW 1166
Db 1111 E--NFKVLSLEGLKLDNLN---LEKKLSYLSGLSHHLLIA-ELKGVINKNY----- 1157
QY 1167 YITSLNPTDFADQCLHEFTHEGDKNLGNISWRITKTLSDSGFFTSVAVQFQD----- 1221
Db 1158 ---TGNSPS-----ENNTDVNNALE---SYK-KFLPEGTDVAVVSESGSDTLEQS 1201
QY 1222 -PNQPGVLYNHDNEYLYHYODDWYILSKIEKPEDYIFVYVYGRNDA-WDGYGGVAVYTR 1279
Db 1202 QPKKPA-----STHVGAESNTITTSQNVDDVDVLIPIFGESEEDYDLGQVY--TG 1253
QY 1280 SSVLPNSIIPLEKAASIGRDEFTRTDNTCGPEPALVERIEKTVBERGIIVKEVEE 1339
Db 1254 EAVTPSVI-----DN-----ILSKIE---NEYEVLYLKPLAG 1282
QY 1340 IEVEVEKEVKVGRTEMTLQRLAEGFNLKQDEENFVRLSKSEMEFLD 1389
Db 1283 VYRSLKQLE---NNVMTFNVNVDIILNSRKNREKNFKNVLESDLIPYKD 1329
RESULT 13
ID MSP1_PLAFW STANDARD; PRT; 1639 AA.
AC P04933;
DT 13-AUG-1987 (Rel. 05, Created)
* DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (P195).
DN MSP-1
* OS Plasmodium falciparum (isolate Wellcome).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5848;
RN [1]

SEQUENCE FROM N.A.
EX MEDLINE=86014355; PubMed=2995820;
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA Nicholls S.C., Hillman Y., Tizard M.L.V., Schwarz R.T.,
RA Freeman R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites.";
RL Nature 317:270-273(1985).
RN [2].
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X02919; CAA26676.1; -
CC PIR; A24594; A24594.
CC InterPro; IPR000561; -
CC Pfam; PF000008; EGF; 1.
KW Malaria; Merozoite; Polypeptide; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 764 764 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 768 768 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 783 783 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 844 844 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 920 920 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;

Query Match 1.88; Score 134; DB 1; Length 1639;

Best Local Similarity 18.0%; Pred. No. 6.7;
Matches 246; Conservative 207; Mismatches 473; Indels 444; Gaps 70;

QY 184 VQFNECAVSRKKVPRKSDVGEFPVPRDRAVQVNFNMKDFSGKWTITSLNPTDFADFC 243
Db 249 IENINELIEESKKTIDK-----NKNATKEEKKLVQAQYDLS-----IVNK 290
QY 244 QLHEPHMENDKLVLGNTLWRIKTLGGPPTRSVQTF--VQDP-----DLPGLYNHDNE 295
Db 291 QLEEAH-----NLISVLEKRIKIDTLKKNENIKELLDKINEKNPPPPANGNTPTLLDKKK 346
QY 296 FLHYODDWYILSSQIENKPDY-----IFVYVYGRNDWDGYSVYTRSPITLPSII 349
Db 347 TEEHEKEKEIAKTIKFNIDSLFTDPLELEYLREKKNID--ISAKVETKESTEPNE-Y 403
QY 350 PNLQKAASVGRDNFNFTTDSGCEPPLVERLEKTAEEGKLLI-----KEAVEIEEE 404
Db 404 PNGVTYPLSY-NDINNALNELNSFG---DLINPFDYTKPSKNIVTDNERKKFNEIKEK 459
QY 405 VKEVEKEVQRTDTEMTLQRLGFGFELQODEENFVRLSKSEKEILNEL-----QMEAT 457
Db 460 IKIEKKKI-----ESDKSKSYEDRSKSLNDITKEYEKLILLNEYSKFNNNIDLT 507

QY 458 EYEKLFGR--ALPIRKLRLAPHSNFIANHETIKYVGVSKLPCHKRPFGWEDY-FGSI 514
Db 508 NFEKMGKRYSYKVEKL-----THNTTASYENSKHNL-EKLTALKY---MEDYSLRNI 558
QY 515 VVAKICSSRRIPRYPRKSPRICCGLDRLGLQFSGHNLSPAHISINONPKNGSGC--- 571
Db 559 VVEKEL-----KYYK-----NLISKIENEI---ETLVENIKKDEQLPEK 595
QY 572 -----KFPKVALMWKVGOFKATAIPIILSVASKADAVDALKTCTCLLKECRLE- 624
Db 596 KITDENKPEKILEV-----SDIVKVOVKVLLM-----NKIDELKTKTOLILKNVELKH 645
QY 625 -----LAKCISN-----PACAA-----NVACLOTQCNRRDET 651
Db 646 NIHVNSYKQENKQBPYLIIVLKEIDKLVFMPKVESLINEEKKNIKTEGSDNSEPST 705
QY 652 EQOI-----KCGDLFNSVDFNECAVSRKKCVPRKSDVGDFFVPDPVSVLQV-- 699
Db 706 EGEITGQATTKPGQAGSALEG---DSVQAQAQEQKQAP-----PVPVPEAKAQP 756
QY 700 -----KFDKDFSGKWFITRGLNPTFDADCOLHEFTEENKLVGNLSWRI 745
Db 757 TPPAPVNNKTENVSKLDYLE-----KLYEFL-----NTSYIC 788
QY 746 RTPDGGFFTRSAVQKVFQDPKPYGILYNHD--NEYLLYQDDWYILSSKVEN-----SPED 798
Db 789 H-----KY--ILVSHSTMNEKILQ---YKITEESKJSSCDPLD 824
QY 799 YIF-----VYKGRNDAMDYGGSV-----LYTRSAVL-----PESTIPELQAAQ 839
Db 825 LLFNIQNNIPVY-----SMFDSLNSLSQLFMEIYKEMVNCNLYKLKDNKIKNLLLEAK 880
QY 840 KVGDFNTFIKTDNCGPEP--PIVERLEKKVBEGET-----IKEYVEE 883
Db 881 KV-----STSVKTLSSSMOPLSLTPQDKPEVSANDDTSHSTNLNLSKLLENILSLGKNK 936
QY 884 -----EEVEKVRDKVETFLSKLFEKGLQDEENFLRELSKEEMVDLGLK 931
Db 937 NIYQELIGOKSENFYEKILKSDTFNSESFTNFKVSKADDIINLNDSEKRR-----K 989
QY 932 MEATEVEKLGFRALPIKRLMAVATHCTSPCHDIRFFSSDDGIGRIGITRKINGTFL 991
Db 990 LE-EDINKL-----KTKLQSLSDLYNKNYKLLERLFDKKTGVKYMQIKKLT---LL 1038
QY 992 KILPIQADLRTGGRSSRP---LSAFRSFGSKGIFDIPLPSKNLKLKELTAPLLKLIV 1048
Db 1039 K-----EQLESKLSLNNPKHVLQNFVFNK-----KKEAE----- 1070
QY 1049 GVLACALFIVPSADAVDALKTACCLLGCGR--TELAKCIANPACAAANVACLOTQCNRRPDE 1106
Db 1071 -----IAETENTLENTKILLKHYKLVKYNGESSPLKTUSESIQTEDNYASL 1119
QY 1107 TECOIKGDLFNSVDFNECAVSRKKCVPRKSDLGFFPAPDPSPVLQVNFISDFNGKW 1166
Db 1120 E--NFKVLSKLEGLKDNLN---LEKKLSYLSGLHLHIA-ELKEVKNKNY----- 1166
QY 1167 YITSGLNPTFADCOLHEFTEGDKNLGNISWRKLTDSGFTFRSAVOKFVOD----- 1221
Db 1167 ---TGNSPS-----ENNTDVNNALE---SYK-KFLPEGTDVATVVSSESGSDTLEQS 1210
QY 1222 -PNOPGVLYNHNDNEYLYQDDWYILSSKIENKPEDEYFVYVYGRNDA-WDGYGAVVYTR 1279
Db 1211 QPKPA-----STHVGAESENTTITSONVDDEVDVLIIVPIGESEEDYDLGQVY--TG 1262
QY 1280 SSVLPNSIIEPELKAASIGRDFSTFIRTDNTCGPEPALVERIEKTEVEGERIIVKEVEE 1339
Db 1263 EAVTPSVI-----DN-----ILSKIE---NEYEVLYLKLPLAG 1291
QY 1340 IEEVEVEKVGTEMTLFLQRLAEGNELKQDEENFVRLSKSEEMFELD 1389
Db 1292 VYRSLKOLE---NNVMTFNVNVKDIILNSRFRNKRNFKNVLESDLIPYKD 1338

RESULT 14

ID MYSB_MESAU STANDARD; PRT; 1934 AA.
AC P13540; Q60540;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
GN MYH7.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F1B; TISSUE=Liver;
RX MEDLINE=95115033; PubMed=7815459;
RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;
RT "Characterization and nucleotide sequence of the cardiac alpha-myosin
heavy chain gene from Syrian hamster.";
RL J. Mol. Cell. Cardiol. 26:1155-1165(1994).
RN [2]
RP SEQUENCE OF 962-1935 FROM N.A.
RX MEDLINE=88247788; PubMed=3380703;
RA Jandreski M.A., Sole M.J., Liew C.-C.;
RT "Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin
heavy chain.";
RL Nucleic Acids Res. 16:4737-4737(1988).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LMH) AND 1 HEAVY MEROMYOSIN (HMH). IT CAN LATER BE
SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
SUBFRAGMENT (S2).
CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; L12104; AAA62313.1; -;
DR EMBL; X07273; CAA30256.1; -;
DR PIR; A28298; A28298.
DR HSP; P08799; IAMD.
DR InterPro; IPR000048; -;
DR InterPro; IPR001609; -;
DR InterPro; IPR002928; -;
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PROSITE; PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 838 GLOBULAR HEAD (S1).
FT DOMAIN 839 1934 RODLIKE TAIL (S2 AND LMH DOMAINS).
FT DOMAIN 779 801 COILED COIL (POTENTIAL).
FT DOMAIN 839 1934 COILED COIL (POTENTIAL).
FT NP_BIND 177 184 ATP.
FT DOMAIN 654 676 ACTIN-BINDING.
FT DOMAIN 756 770 ACTIN-BINDING.

```

FT MOD_RES 128 128 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 966 966 D -> E (IN REF. 2).
FT CONFLICT 978 978 T -> TE (IN REF. 2).
FT CONFLICT 986 986 E -> Q (IN REF. 2).
FT CONFLICT 1008 1014 DLQARED -> ALEARKT (IN REF. 2).
FT CONFLICT 1057 1057 D -> Y (IN REF. 2).
FT CONFLICT 1060 1060 L -> V (IN REF. 2).
FT CONFLICT 1095 1095 D -> N (IN REF. 2).
FT CONFLICT 1217 1217 E -> D (IN REF. 2).
FT CONFLICT 1271 1271 D -> N (IN REF. 2).
FT CONFLICT 1327 1327 T -> A (IN REF. 2).
FT CONFLICT 1358 1358 C -> R (IN REF. 2).
FT CONFLICT 1504 1504 L -> V (IN REF. 2).
FT CONFLICT 1537 1537 M -> L (IN REF. 2).
FT CONFLICT 1556 1556 N -> K (IN REF. 2).
SQ SEQUENCE 1934 AA; 222928 MW; FDBAC58310B0B57D CRC64;

Query Match 1.8%; Score 133.5; DB 1; Length 1934;
Best Local Similarity 17.0%; Pred. No. 8.9; 525; Indels 493; Gaps 63;
Matches 251; Conservative 211; Mismatches 211;

QY 188 NECAVSRKKCVPRKSDGEEFVPRDN-----AVQNFNMKDFSGKWIITSGLNPT- 237
DB NSADLLKGMCHPRVKGVNEVTKQNVQVQVYAIGALAKSVYEKMFN--WMVTR-INATL 446
QY 238 -----FDAFDQOLHE---FHMENDKLVGNLTWRIKTLDDGFFTSVAQ 277
DB 447 ETKOPROYFIGVLDIAGFEIFDFNSFQPCINFTEKLOOFFNHHMFVLEQEYEEKREGIE 506
QY 278 -TFVQ-----DPDLPGLAG--NHDNEFLHYQDDW 303
DB 507 WTFDFDGMQACIDLIEKPMRIMSILIEECMPKATDMTFKAKLYDNHLGKSNFQKPR 566
QY 304 YILSSQENKPDYIFYVYGRND-----AWDGYGGSVIYTRSTPLPESIIPIPNQKAA-KS 358
DB 567 NVKQKQEAH-----PSLVHYAGTVYNTLGN-----LQNKQPLNATVGLYQKSSKL 615
QY 359 VGRDFNFIITDSCGPEPLPVERLEKTAEGEKLLIKEAVEI--EEVEKEVEKVRDTE 416
DB 616 LSNLFANY-----AGADAP-VDKKGKAKKGSF---QIVSVLHRENKLNKLTNLRST- 664
QY 417 MTLFQRLLEGFKELQDDENFVRLSKEEKE-----ILNLEOMEAT-EVEKLFGR 466
DB 665 -----HPHFVRCIIPNETKSPGVMDNPLVMHQLRCNGVLEGIRICRG 707
QY 467 LP-----IRKRLMALPHSNLANHETIKYVVGSKLPCHKRFSGWEDYFGSIV 515
DB 708 FPNRILYGDFFQRYRILNPAIPEGQFIDSRKGAELKLLSLDDIDHNOYKFGHTKVFFKAG 767
QY 516 VAKICSSRRIPRYKSPRICGDLRGLQFSGKHNLSPAHS---INONVPKNGSGCK 572
DB 768 ILGLLEMRDRSLRIITRI--QAQSRGLLSRMEFKLLERRDSSLVIQWNI-RAFMGVK 824
QY 573 -FP-----KOVALMWKWKQGFATAIYVAIFILSVASKADAVDAKLT 613
DB 825 NWPWMKLYFKIKPLKSAETEKEMATMK-BEFGR-----VKDALEK 864
QY 614 CTCULLKECRLELAKISNPACAAVACLOTNNRPDETECIQKCGDLFENS-----VDE 668
DB 865 SEARKKELEKMWLLQBP-----KNDLQLOVQAEQDNLADEERCDQLKKIQLEAKVKE 920
QY 669 FNE-----CAVSRK---KCVPRKSDGDFVPDPSPVLVQKDFMKDFSGKWFIT 713
DB 921 MTERLEDEENNAELTAKKRKLDEBCESELKRDIDDLTLAKVEKDK----- 967
QY 714 RGLNPTFAFCQLHEFTEENKLVGNLSWRIRTPDGGFFTRSAVQKVFQDPKYPGLIYN 773
DB 968 -----HATENK-VKNLTEEMAGLD-----ETIAKLTKKKALQBAHQ 1003
QY 774 HDNEYLLYQDDWY--ILSSKVE-----NSPEDYIFVYVYKGRND---AWDGYGGSVLVTR 823

```

```

DB 1004 QALDQLQAEEDKVNLTITKSKVKLEQVDDLEGSLEQEKVVRMDLERAKRKLQEGDKLTQE 1063
QY 824 AVLPEIPIPELOTAQKVG-----RDF-----NTFIKDTNCTGPE-----PPLVERL 865
DB 1064 SIM-----DLENDKQQLDEKLLKKKDFELNALNARIIDEQALGSQLQKKLKEIQARIEEL 1117
QY 866 EKKVEGERT-----IIEVEEIEEVEK----- 889
DB 1118 EEL-BAERTARAKVEKRLSDLSRELEIEERLEEAGGATSVQIEMNKKREAEFQKMRD 1176
QY 890 -----VRKEVTLFSKLEFEGFKELORDENFURELSKEEMVLDGLKMEAT 935
DB 1177 LEAATLQHEATAAALRRKHADSVAELEGOIDNLRQVKOLEKSEKSEFKLEDD-----VTS 1232
QY 936 EVEKLFGRALPIRKLMAVATHCTSPCHDRIRFFSDGIGRLGITTRKIRINGTELLKILP 995
DB 1233 NMEQIITAKANLEKM-----CRT--LEDQNEHRS-----KAETORSVNDL----- 1272
QY 996 PIQSADLRTTGTGRSRPL---SAFRSGFSKGIIDFIVPLPSKNELKELTAPLLLLKLGVLA 1052
DB 1273 TSQRAKLOTENGELSRLQDEKEALISQLTRG-----KLTYTQQLLEDLKRQL----- 1318
QY 1053 CAFILVPSADAVDAKLTCACLLKGCRIELAKCIANPACAAVACLOTNNRPDETE--CQ 1110
DB 1319 -----EEVYKAKNTLAHALQSAHND-----CDLLREQYEEETEAKAE 1355
QY 1111 IKC-----GDLFENSVVDEFNECAVSRKKCVPRKSDGELGEPFAPDPSPVLVQNFN 1158
DB 1356 LOCVLISKANSEVAQWRTKYETDAIQTETEELKAKLAQRLQDAEE----- 1401
QY 1159 ISDFNGKWIYISGLNPTDADFQQLHEFHTEGDNKLVGNISWRITKLDGFFTSVAQKF 1218
DB 1402 -----AVEAVNAKCSSEKTKHRLQNEIEDLMDV-----VERSNAAAA 1439
QY 1219 VODPNQPGVLVNHNDHYLYHODDWYILSSKIENKPEYIFVYVYGRNDAMDGYGVAVYT 1278
DB 1440 ALDKKQ-----RNFDKLLAEWKQYEESELESSEKE-----ARSLST 1478
QY 1279 RSSVLNPIIPELE-----KAASIGRDFSTFIKDTNCTGPEPALVERIEKTVEEGERI 1332
DB 1479 ELFLKLNAYEESLEHLETFKRENKNIQBEISDLTQLOSTGKSIHELEKIRKQLEAEKME 1538
QY 1333 IVKEVEEIEEVEKEVEKVGRTMTLFO-----RLAEGFENLKODEENFVRLS----- 1381
DB 1539 LOSALEEAEASLEHEEGNILLRAQLEFNQIKAEIERKLEKDEEMEQAKRNHLRVVDSLOT 1598
QY 1382 -----KEEME-FLDEIKMEASEVEKLFGKA 1405
DB 1599 SLDAETRSRNEALRVKKMKGDLNEMEIQLSHANRMAAEA 1638

RESULT 15
ID LEF_BACAN STANDARD; PRT; 809 AA.
AC P15917;
DT 01-APR-1990 (Rel. 14; Created)
DT 01-APR-1990 (Rel. 14; Last sequence update)
DT 15-JUL-1998 (Rel. 36; Last annotation update)
DE LETHAL FACTOR PRECURSOR (EC 3.4.24.-) (LF).
GN LEF.
OS Bacillus anthracis.
OG Plasmid pX01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-49.
RX MEDLINE=90034185; PubMed=2509294;
RA Bragg T.S., Robertson D.L.;
RT "Nucleotide sequence and analysis of the lethal factor gene (lef)
from Bacillus anthracis.";
RL Gene 81:45-54 (1989).

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 05:03:46 ; Search time 119.14 Seconds
(without alignments)
1568.027 Million cell updates/sec

Title: US-09-075-375A-2

Perfect score: 7495

Sequence: 1 MALSLHTVFLCKEALNYA.....MEASEVEKLGKALPIRKVR 1412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB-seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_16:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phase:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.unclassified:*
- 13: sp.invertebrate:*
- 14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	2557	34.1	478	10	Q40593
2	2508	33.5	473	10	Q40251
3	2430	32.4	462	10	Q39249
4	1604.5	21.4	472	10	Q39249
5	1300	17.3	244	10	Q39249
6	1219.5	16.3	289	10	Q39249
7	210	2.8	522	10	Q39249
8	194.5	2.6	269	5	Q26223
9	161.5	2.2	2771	5	Q26216
10	156	2.1	1365	2	Q49525
11	155.5	2.1	1933	13	Q39237
12	151	2.0	1055	10	Q39230
13	151	2.0	1939	5	Q39230
14	150.5	2.0	1332	4	Q39230
15	148.5	2.0	2473	11	Q39230
16	147.5	2.0	1786	5	Q39230
17	146.5	2.0	1387	5	Q39230
18	146	1.9	1935	4	Q39230
19	146	1.9	1935	4	Q39230

20	146	1.9	2748	3	Q03767
21	145.5	1.9	1116	5	Q03767
22	144.5	1.9	839	5	Q26024
23	144	1.9	2166	2	O51465
24	144	1.9	3899	4	O396Y2
25	143.5	1.9	1558	5	O396Y2
26	143	1.9	1978	5	Q9W0M1
27	142.5	1.9	880	1	Q9U2C8
28	142.5	1.9	1199	5	P91349
29	142.5	1.9	1819	2	Q92LV0
30	142.5	1.9	5105	5	O61201
31	141.5	1.9	1935	6	O9GK81
32	141	1.9	1002	2	O66583
33	140.5	1.9	1302	2	O49547
34	140.5	1.9	1930	13	Q9DGD5
35	140	1.9	800	1	Q59066
36	140	1.9	3595	4	Q9UQH3
37	139	1.9	2712	10	Q9SB74
38	139	1.9	3911	4	Q99996
39	138.5	1.8	1327	4	Q9Y2L2
40	138	1.8	1057	10	Q9Y2L2
41	138	1.8	2867	5	Q9N2M3
42	137.5	1.8	3908	4	Q9UQ04
43	137	1.8	1088	4	O60772
44	137	1.8	1109	6	O00756
45	137	1.8	1676	10	O23332

ALIGNMENTS

RESULT 1

Q40593 ID AC Q40593; PRELIMINARY; PRT; 478 AA.
 DT 01-NOV-1996 (TREMREL. 01, Created)
 DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
 DT 01-MAY-2000 (TREMREL. 13, Last annotation update)
 DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
 GN TVDEL.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
 OC Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=XANTHI; TISSUE=LEAF;
 RA BUGOS R.C., Yamamoto H.Y.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U34817; AAC50031.1; -
 DR Mende; 9222; Nicta:Vdel;9222.
 DR InterPro; IPR000566; -
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
 DR TRANSIT. 134 POTENTIAL.
 FT CHAIN 135 478 POTENTIAL.
 SQ SEQUENCE 478 AA; 54561 MW; 0967DF4547D7809D CRC64;

Query Match 34.1%; Score 2557; DB 10; Length 478;
 Best Local Similarity 100.0%; Pred. No. 6.9e+138;
 Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	474	MALAPHSNLANHETIKYVGSKLPGHKRFSWGWDYFGSIWAKICSSRRIPYFRKSP	533
Db	1	MALAPHSNLANHETIKYVGSKLPGHKRFSWGWDYFGSIWAKICSSRRIPYFRKSP	60
Qy	534	RICGGLDSRGQLFSGHKHNLSPAHISINQNPVKGNSCKFPKDVAMVWEKWFQAKTAI	593
Db	61	RICGGLDSRGQLFSGHKHNLSPAHISINQNPVKGNSCKFPKDVAMVWEKWFQAKTAI	120
Qy	594	VAFILSVASKADAVDALKTCTCLLKECRLELAKICISNPACAAVACLTQTCNNRDPDETC	653

```

Db 121 VAFILSVASKADAVDALKTCTCLLKECLELAKCISNPACANVACLOTCNRRPDETEC 180
QY 654 QIKCGDLFNSVVDNECAVSRKCVPRKSDVGDGPPVDPSPVLQKDFMDKDFSGKWFIT 713
Db 181 QIKCGDLFNSVVDNECAVSRKCVPRKSDVGDGPPVDPSPVLQKDFMDKDFSGKWFIT 240
QY 714 RGLNPTFADFQCLHBFHTEENKLVNLSWRIRTPDGGFFTSRQVQDPKYPGILYN 773
Db 241 RGLNPTFADFQCLHBFHTEENKLVNLSWRIRTPDGGFFTSRQVQDPKYPGILYN 300
QY 774 HDNEYLLYQDDWILSKVNSPEDYIFVYKGRNDADWDGYSGLVYTRSAVLPESIIPE 833
Db 301 HDNEYLLYQDDWILSKVNSPEDYIFVYKGRNDADWDGYSGLVYTRSAVLPESIIPE 360
QY 834 LQTAQKVGGRDNFTIKTDNCTGPEPLVERLEKKVEEGERTIIKEVEIEEVEKVRDK 893
Db 361 LQTAQKVGGRDNFTIKTDNCTGPEPLVERLEKKVEEGERTIIKEVEIEEVEKVRDK 420
QY 894 EVTLFSKLFEGFELQDEENFLRELKSKEEMDVLDGLKMEATEVEKLFGRALPIRL 950
Db 421 EVTLFSKLFEGFELQDEENFLRELKSKEEMDVLDGLKMEATEVEKLFGRALPIRL 477

RESULT 2
Q40251 PRELIMINARY; PRT; 473 AA.
AC Q40251;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN VDEL.
OS Lactuca sativa (Garden lettuce).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
OC euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.
OX NCBI_TaxID=4236;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ROMAINE;
RX MEDLINE=96270536; PubMed=8692813;
RA Bugos R.C., Yamamoto H.Y.;
RT "Molecular cloning of violaxanthin de-epoxidase from romaine lettuce
RT and expression in Escherichia coli.";
RL proc. Natl. Acad. Sci. U.S.A. 93:6320-6325(1996).
DR EMBL; U31462; AAC49373.1; -;
DR Mendel; 8691; Laccsa; Vdel; 8691.
DR InterPro; IPR000566; -;
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW TRANSIT peptide.
FT CHAIN 126 473 VIOLAXANTHIN DE-EPOXIDASE.
SQ SEQUENCE 473 AA; 54447 MW; 1B22522DC2C62699 CRC64;

Query Match 33.5%; Score 2508; DB 10; Length 473;
Best Local Similarity 100.0%; Pred. No. 4.2e-135;
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALSHTVFLCKEALNLYARSPCNERFHRSGQPPTNIMMKIRSNNGYNSFLFTSYK 60
Db 1 MALSHTVFLCKEALNLYARSPCNERFHRSGQPPTNIMMKIRSNNGYNSFLFTSYK 60
QY 61 TSSFSDSHCKDKSQICSDITSFEEIQRFDLKRGMTLILEKQWRQFIOLAIVLVCTEVIV 120
Db 61 TSSFSDSHCKDKSQICSDITSFEEIQRFDLKRGMTLILEKQWRQFIOLAIVLVCTEVIV 120
QY 121 PRVDAVDALTKACLLKCEKRELAKCIANPSCANVACLOTCNRRPDETECQIKCGDLFE 180
Db 121 PRVDAVDALTKACLLKCEKRELAKCIANPSCANVACLOTCNRRPDETECQIKCGDLFE 180
QY 181 NSVVDQNECAVSRKCVPRKSDVGEFPVDRNAVQNFNMKDFSGKWIYTSGLNPTFDA 240

```

```

Db 181 NSVVDQNECAVSRKCVPRKSDVGEFPVDRNAVQNFNMKDFSGKWIYTSGLNPTFDA 240
QY 241 FDCQLHFEHMDKLVNLTWRIKTLDGGFFTSRQVQDPDLPGALYNHNDNEFLHYQ 300
Db 241 FDCQLHFEHMDKLVNLTWRIKTLDGGFFTSRQVQDPDLPGALYNHNDNEFLHYQ 300
QY 301 DDWYTLSSQIENKPDYIFVYGRNDADWDGYSGLVYTRSPITPESIIPLNLOKAAKSVG 360
Db 301 DDWYTLSSQIENKPDYIFVYGRNDADWDGYSGLVYTRSPITPESIIPLNLOKAAKSVG 360
QY 361 RDNFNITDNSCGPEPLVERLEKTAEGEKKLLIKEAVEIEEVEEVEKVRDTMTLF 420
Db 361 RDNFNITDNSCGPEPLVERLEKTAEGEKKLLIKEAVEIEEVEEVEKVRDTMTLF 420
QY 421 ORLEGGFELQDEENFVRELKSKEKELTINLOMEATEVEKLFGRALPIRL 473
Db 421 ORLEGGFELQDEENFVRELKSKEKELTINLOMEATEVEKLFGRALPIRL 473

RESULT 3
Q39249 PRELIMINARY; PRT; 462 AA.
AC Q39249;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN AVDEL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLOMBIA;
RA Bugos R.C., Yamamoto H.Y.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chioi J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Ienz C., Li J., Liu A., Liu J., Liu S., Mukharsy N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U44133; AAC50032.1; -;
DR EMBL; AC003981; AAF99753.1; -;
DR Mendel; 6341; Arath; Vdel; 6341.
DR InterPro; IPR000566; -;
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW SIGNAL.
FT SIGNAL 1 113 POTENTIAL.
FT CHAIN 114 462 VIOLAXANTHIN DE-EPOXIDASE.
SQ SEQUENCE 462 AA; 52017 MW; 5B37B2C1D2D4426B CRC64;

Query Match 32.4%; Score 2430; DB 10; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.1e-130;
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 951 MAVATHCFTSPCHDRIRFSSDDGIGRIGITRKRIKNGITFLKILPPIQSADLRTTGGRSS 1010
Db 1 MAVATHCFTSPCHDRIRFSSDDGIGRIGITRKRIKNGITFLKILPPIQSADLRTTGGRSS 60
QY 1011 RPLSAFRSFGSGIFDVPPLPSKNELKELTAPLLKLVGLVACAFILVPSADAVDALKTC 1070
Db 61 RPLSAFRSFGSGIFDVPPLPSKNELKELTAPLLKLVGLVACAFILVPSADAVDALKTC 120
QY 1071 ACLLKCCRIELAKCIANPSCANVACLOTCNRRPDETECQIKCGDLFNSVVDNECAV 1130

```


QY 502 RFSWGWDYFGSIVVAKICSSRRIPRYFRSPRICCGLDGRLQLFSGHKNLSPARS-I 560
Db 447 -----INTNKNPNSELISQTSKSDSKNSVTDSSNKSDESANTEL 489
QY 561 NONVPKMGSGCKPKDVALMWKMGWQAKTAIVAIFILSVASKADAVDAKTCCTCLLKE 620
Db 490 KQALAKAN-----ADKVAQ-----DNLAKS 509
QY 621 CRLELAKCISNPACANVACLOTNNRPDETECOIKCGDLFENSVDDEFNECAVSRKCV 680
Db 510 IKEQLNNSVSN-----ANTLSAKLTDKDN--TIQAKTE 541
QY 681 PRKSDVGDFPVPDPSVLVQFKDMKDFSGKWFITRGLNPTFDA-----PDCOLHEFHTEENK 736
Db 542 LEKE-----VQADQAISNNTASMSAKSSLDKVAEITKKLETFNKKDEA 598
QY 737 LVGNLSWRITPDGGFTTRSAVKFVQDPKYPGILYHNDNEYLLYQDWWYLSKVENSP 796
Db 589 KFNEKQ-----TRNQIQEFINTK-----NNPN-----YSELISQTSKRDS-- 626
QY 797 EDYIFVYKGRDWDGYSGLVTRSAVLPELIIP-----ELQTAQKVGDRDFTFIK 850
Db 627 -----KNSVTDSSNKSDESANTELQALAKANAKVQADNLAKSIKQLNNSVS 676
QY 851 TDNTCGPEPLVERLEKKVEGERTIIKEVEIEEVEEK-----VRDKEVTLF 898
Db 677 NANT-----LSAKLTDKDNTIOQAQTELEKELOKANQAISNNTASMSAKSSLD 726
QY 899 SKLEGGKELQ-----RDENFURELSKEMDVLGDKMWEATEVEKLFGRALPIRKLMAVAT 955
Db 727 AKVAEITKKLETFNKKDEAKFNEKQTRNQI-----QEFINTNK----- 765
QY 956 HCFTSPCHDRIRFFSDGIGRLGITRKRITNGTFLLLPILQIAD-----LETTGGRS 1009
Db 766 ---NNPNSEL-----ISQTSKSDSKNSVTDSSNKSDESANTELQALAKAKK 814
QY 1010 S-----RPLSAFRSGFSKGIFDVPL-----PSKNEL-KELTAPILLKLVG 1049
Db 815 SSIDNELRPL---KNDLQSKIEEGPIRINTNFSWISSKLETTKKNLAEFLT----- 862
QY 1050 VLACFLVPSADAV-----DALTCACLLKGRICELAKCIANPACANVACLOTGN 1101
Db 863 -----KADAIKNPNSSSKOALKDSSQVQKLGNELKLTITE-----EFGKVETKN 907
QY 1102 NRPDETECOIKCGDLFENSVDDEFNECAVSRKCVPRKSDLEGPAPD-----PSVLVQ 1155
Db 908 SNIGYRLFKLAQAQFNNSVDKLNKNAWEKQTLSSKKQLGNQSTKDYLTQLTSTENSTQ 967
QY 1156 NFNISD--FNGKWYITSGLNPTFDQFCQLEHFEHTEGDNKLVGNISWRITLDSGFETR- 1212
Db 968 ESTIKKIVNIOAHIRNLN-----SOYPLEAD-KLIAN-----MKRGYDKV 1009
QY 1213 --SAVKFVQDPNPGVLYNHNDNEYLYHQDDW-----YILSSKIENKPDYIFVYVGRN 1265
Db 1010 GIESLQKW-QDLMDSDSVLSDVS-----LKDDFNKALRVLVGDYTKNPPVSWFINKNRNS 1064
QY 1266 -DAWDGYGAVVYTRSSVLPSNI-----IPELEKKAKSIGR----- 1300
Db 1065 IENTONRLNLVRENEILLDKADKDKRAEKTIFVDENINSIDQRAKRLKQELNKN 1124
QY 1301 DFSTFIR-----TDNTCGPEPALVERIEKTEVEGERIIVKE-----VEIEEVEEVEK 1348
Db 1125 DLSNFTLNHQNQAKTDITPKISLELKNELNEINOYLLPIIKEKAVKISEIEKN-KKEL 1183
QY 1349 EKVGRTEMTLQRLAEGFNEKQDEENFVRELSKEEMEFLDEIKME 1394
Db 1184 EDIIRSNFYLWE-----KVEINKYISELTNTKQVELSRNSINFE 1220
RESULT 11
Q90337
ID Q90337 PRELIMINARY; PRT; 1933 AA.
AC Q90337;

DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE MYOSIN HEAVY CHAIN.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
OX NCBI_TaxID=7962;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FAST SKELETAL MUSCLE;
RX MEDLINE=97352533; PubMed=9208928;
RA Hirayama Y., Watabe S.;
RT "Structural differences in the crossbridge head of temperature-
associated myosin subfragment-1 isoforms from carp fast skeletal
muscle.";
RL Eur. J. Biochem. 246:380-387(1997).
RN [2]
RP SEQUENCE OF 738-1933 FROM N.A.
RC TISSUE=FAST SKELETAL MUSCLE;
RX MEDLINE=97176447; PubMed=9023993;
RA Imai J., Hirayama Y., Kikuchi K., Kakinuma M., Watabe S.;
RT "cDNA cloning of myosin heavy chain isoforms from carp fast skeletal
muscle and their gene expression associated with temperature
acclimation.";
RL J. Exp. Biol. 200:27-34(1997).
DR EMBL; D89990; BAA22067.1; -;
DR EMBL; D50474; BAA09067.1; -;
DR HSSP; P08799; LLVK.
DR InterPro; IPR000048; -;
DR InterPro; IPR001609; -;
DR InterPro; IPR002928; -;
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; -; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
KW Myosin.
SQ SEQUENCE 1933 AA; 221092 MW; CDF0CBAA475530F5 CRC64;
Query Match 2.1%; Score 156; DB 13; Length 1933;
Best Local Similarity 17.0%; Pred. No. 2.1;
Matches 203; Conservative 181; Mismatches 454; Indels 356; Gaps 42;
QY 379 LVRLERKTAEGEGKLLIKEAVEIEEVEKEVEKVRDTEMTLFORLLEGFKE-----L 430
Db 606 VVQLYQKSAKLVALLVAVPAEAAAGKGGKGGKGS----FQTVSAVRENGLKMTNL 661
QY 431 QQDEENFVRELSKEEKE-----ILNELOWE-ATEVEKLFGRALPTR----- 470
Db 662 RSTPHFVRLIPNESKTPGLMENFLVIHQRCNGVVEGIRICTKGFPSRIHYGDFKQRY 721
QY 471 -KLRMALAPHSNLANHETIKYVYVGSKLPGHFRFSGHWEDYF---GSIIVAKICSSRRIP 526
Db 722 KVLNASVPEGQFIDNKKATELLGSDIDVHNQYKFGHTKVFVKAGLLGTLEEMRDEKLS 781
QY 527 RYFRKSPRCCGLDLSGLQFLSHGKHNLSPAHSINONVPKNGSGCKFKPDVALMWYKMG 586
Db 782 HLVMTQALARGVVR--KEFKVMERRRAIYSIQYINIRSPMNKHWV---WMKYVFKIK 836
QY 587 QFAKTAIVAIFAIFILSVASKADAVDAKTCCTCLLKECRLELAKCISNPACANVACLOTGN 646
Db 837 PLLKTAESE---KEMASMKENFEKKEDITKALAKKLEEKWVSLVQEKNDLLQVTSE 893
QY 647 RPDTECOIKCGDLFENS-----VDEFNECAVSRK-----CVRPKSDV 686
Db 894 SENLSDAEERCEGLIKSKIQLEKLEKLETTTERLEDEBEINAEITAKKRLDEKSELKKDI 953
QY 687 GDFPVPDPSVLVQKFDMDKDFSGKWFITRGLNPTFDQFCQLEHFEHTEENKLVGNLSWRIR 746

```
Db 954 DDELTLAKVEKEK-----HATENK-VKNLTEEMA 982
QY 747 TPGGGFTTSVAVQKFDQPKYPCILYNHNDNEVLLYQDDWYILSSKVENSPEDYIFVYK 806
Db 983 SQD-----ESTAKLTKEKKAQLOHAHQOQLDLDQAEKVNTLTAKTKLEQOV----- 1030
QY 807 RNDWDGYSGLVYTRS-----AVLPESIIIP-----ELQTAQKVGK----- 843
Db 1031 -----DDLEGSLEOEKRLMDLVRKLEGLDKLAQESIMDLENEKQSDDEKIKKDFE 1085
QY 844 --DPTNFIKNTGCP-----PPLVERLEK-----KVEGERTIIKEVEEI 883
Db 1086 ISOFLSKIEDQSLGAQLOKIKELQARIEELEETEAEARSARAKVEKORADLSRELEI 1145
QY 884 EEEVEK-----VROKDEVTLFSKLEPG 904
Db 1146 SELEEAGGATAAQIEMNKKREAFQKMRDLLEESTLQHEATAAALRKQADSVALEGQ 1205
QY 905 FRELQDEENFLRELSKEEMDVLGLKMEATEVEKILFGRALPIRKLMAVATHCFTSPCHD 964
Db 1206 IDNLQRIKIKLEKESEYKMEI--DOLSSNNEAVAKAGN---LEKN-----CRT--LED 1253
QY 965 RIRFF--SSDDGIGRLGITKRINGTFLKILPPIQSADLRTGGSSR-----P 1012
Db 1254 OLSEIKAKSDENSRQNDMN-----AQRARIQTENGFSFQLEKEEALVSQ 1299
QY 1013 LSAPRSFGSKGIPDIIVPLPSKNELKELTAPLLKLVGLACAFILVPSADAVALKTCAC 1072
Db 1300 LTRGKAFTQOIEDL-----KRHVEE----- 1320
QY 1073 LLKGRICELAKTIANPACANVAC--LQTCNNRPDTECOIKCG-----DLFE 1118
Db 1321 -----EVKAKNALAHAVQARHDCDLRLRQYEEQEAELQRLGMSKANSEVAQWRKRYE 1375
QY 1119 NSVDFEFNCAVSRKCVPRKSDGLGFPAPDPSPVLVQNFNISDFNGKWYITSGLNPTEDA 1178
Db 1376 TDAIORTEELSEKSKLAQLQDAE-----SIEAVNSK---CASLEKTQR 1419
QY 1179 FDCQLHEFTGPD--NKLVGNIISWRIKTLDSGFTTSVAVQKFDQPNQPGVLYNHNDNEYL 1236
Db 1420 LQSEVEDLMIDGERANALANLDKQKNFD-----KVLADWKQK---YEESQAE 1466
QY 1237 H-YQDDWYILSSKIENKPEDYIFVYVGRNDWDGYSGLVYTRSVPNSIIPLEKAA 1295
Db 1467 EAAQKEARSLSLTFKKNYS-----EALDH-----LETIKREN 1501
QY 1296 KSGIRDFSTFIRTDNTCGPEPALVERIEKTVEEGERIIVKEVEIEEVEKEVKGRT 1355
Db 1502 KNLQEQEISDLSEOLGETGKSIHIEKAKKTVESEKAEIQTALAEAGTLEHEESKILRVQ 1561
QY 1356 MTLFQ-----RLASGFNELKQDENFVRELSEKMEFLEIKM--EASEVEK 1400
Db 1562 LELNQVKEIDRLKLADEMEFOIKRNSQRLDSQMSQTLDSVRNDRALRVKK 1615

RESULT 12
O29230 PRELIMINARY; PRT; 886 AA.
AC O29230;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PURINE NTPASE, PUTATIVE.
GN AF1032.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OX Archaeoglobus.
RN NCBI_TaxID=2234;
RP [1]
RC SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RC MEDLINE=98049343; PubMed=9389475;
RX
```

```
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervage A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., Mckenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RA "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
DR EMBL; AE001032; AAB90211.1; -.
DR TIGR; AF1032; -.
DR InterPro; IPR001238; -.
DR InterPro; IPR002017; -.
DR InterPro; IPR003439; -.
DR Pfam; PF00470; Recf; 1.
KW Hypothetical protein.
SQ SEQUENCE 886 AA; 103633 MW; D35641D499AAB58 CRC64;
```

Query Match 2.1%; Score 155.5; DB 1; Length 886;
Best Local Similarity 20.6%; Pred. No. 0.78;
Matches 137; Conservative 85; Mismatches 221; Indels 221; Gaps 27;

```
QY 797 EDYIFVYVYKGNDAWDCYGGSV--LYTRSALVPESIIPELQTAQKVGRODNFTFIKTD-N 853
Db 152 EDY-----ENAWKNLGAIVMLEREKELFEFSQEQIKRQEEK-----KAEIE 197
QY 854 TCGPEPPLVERLEKKVEEGERTI---IKEVEIEEVEEVEKVRDEKVEFLSKLFGFELQR 910
Db 198 RISEIKSIESRLSEKLESEVRNLSRLKEHKSRLSEKRSLEKQESSVLQEV-----R 249
QY 911 DEENFLRELSEKEMDV--LDGLKMEATEVEKLFGRALPIRKLMAVATHCFTSPCHDRIR 967
Db 250 GLEEKLEKQLKEKVERIEDLEKAKEKVEL-----RKPAERY- 289
QY 968 FFSDDGIGRLGITRKRINGTFLKILPPIQSA--DLRTTGGSSRPLSAFRSGFSGKIF 1025
Db 290 -----SILEKLLSEINQALRDVEKREGDLTREAAQIAQLKAAE 329
QY 1026 DIVPLPSKNELKELTAPL-----LLKLVGLVACAFILVPSADAVDAKTC 1070
Db 330 D-----NSKLEETIKRIELELELEFEKSHRLE-----TLKPKMDRMQGIK-- 372
QY 1071 ACLLKGRICELAKIANPACANVACIQTNNRPDTECOIKGDLFPENSVDPEFNECAV 1130
Db 373 -----AKLEKNLTPDKVE---KMYDLSLKAKEEE-KEITE 404
QY 1131 SRKCVPRKSDGLGFPAPDPSPVLVQNFNISDFNGKWYITSGLNPTFDAPCQLHEPTEG 1190
Db 405 KLKLIKAKKSLKTRGAQLKAAVEE-----LKSARTCPVCGRELEDEHRKN 451
QY 1191 -----DNKLVGNISWRIKTLDSGFTTSVAVQKFDQPNQPGVLYN-- 1230
Db 452 IMAEYTRMKRIAEELAKADEIEKKLEKLEKVEKALEKQFTVLKYRQWVDELKALENEL 511
QY 1231 --HDNEYLHYQDDWYILSSKIENKPEDYIFVYVGRNDWDGYSGLVYTRSVPNSI 1288
Db 512 SSHDAE-----KLSAESEE-----YRKVKERLDGLRGOOKILLSSA--SRI 550
QY 1289 PELEKAASIGROFSTFIRTDNTCGPEPALVERIEKTVEEGERIIVKEVEIEEVEKEV 1348
Db 551 KELASSUREI-----EEL-KNVESERGELHHRKIREGFESELEEREV 593
QY 1349 EKVGRTEMTLFQRLAEGFN--ELKQDEENFVRELSEKMEFLEIKMEASEVEKILFGKA 1405
Db 594 -----QSLRPFPYKNWLEKDAESRLSEKRRR-KLEDEITSEATAKLEANGKA 641
QY 1406 LPIR 1409
```

```

Db 642 EEIR 645

RESULT 13
Q9S722 PRELIMINARY; PRT; 1055 AA.
AC Q9S722;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE SMC-LIKE PROTEIN.
GN MIM.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=99380167; PubMed=1049416;
RA Mengiste T., Revenkova E., Bechtold N., Paszkowski J.;
RT "An SMC-like protein is required for efficient homologous
recombination in arabidopsis.";
RL EMBO J. 18:4505-4512(1999).
DR ENBL; AFI20933; AAD54770.1; -
DR ENBL; AFI20932; AAD54769.1; -
DR InterPro; IPR003439; -
SQ SEQUENCE 1055 AA; 121349 MW; E9F0C0427FB602E4 CRC64;

Query Match 2.0%; Score 151; DB 10; Length 1055;
Best Local Similarity 19.3%; Pred. No. 1.8;
Matches 182; Conservative 119; Mismatches 306; Indels 334; Gaps 43;

QY 128 ALKTCACLLKECRIE-----LAKTANPSCAAVACLOPCNNRPDETECOIKCGDLFEN 181
Db 57 AILTALCIAFGARGQRAATLKDFIKTCSYAVVQVEMKNSGEDAFKSEIYGVIIIE 116
QY 182 SVVDQNECAVSR-----KKCVPRKSDVGEPPVDRNAVONFNM-----KDFSGK 227
Db 117 RRTESATATVLKDYLGKYSKRDRE-----LVEHFNIDVENPCVMSODKAGS 168
QY 228 WYITSGLNPTFDAPDCOLHEFHENDKLVGNLTWRIKTLGDGFFTRSAVOTFTVQDDPLPG 287
Db 169 SYI-----LECK-----GN-----SSSFLRLNLIQ---QVNDLIQ 194
QY 288 ALYNHNDNEFLHYQDDWYILSSQIEN--KPDYIFVYVYGRNDAMDGYGGSVIYTRSPITLP 345
Db 195 SIYEHLTAKATAVD-----ELENTIKPIEKEISELRGK-----IKNMEQV 234
QY 346 ESIIIPNLQAAK-----SVGRDFNNFTTNSCGPEPPVERLEKTAEGEKKLLIKE 397
Db 235 EEIAQRLOQLKKLWNSWYDVGRQ-----LQEQTEKIVKLKERIPTCQ 278
QY 398 AVETEVEVEVEKVRDT-----EMTLFORLLEGEKE-----LQOD 433
Db 279 A-KIDWELGK-VESLRDTLTKKAQVACLMDSTAMKRETESPHQSAKTAVRKIALQEE 336
QY 434 -----BENFVRLSKKEKEILNLOMEATEVEKLFGRALP 468
Db 337 FNHKCNVQKIKDVRRLERQVGDINEQTMKNQAQSEIEEKLKYLEREVEK----- 389
QY 469 IRKLRLMALAPHSN-----FLANHETIKYVGVSKLPDGHKRGKSWGWEDY 510
Db 390 VETLRSLKEEENCFLKAFEGRKKMEHIEDMKNHQKQRFITTSINDLKKHQTAKVTA 449
QY 511 FGSITWAKICSS--RRIPRYFRKSPRICCGDLDSRLQLFSGHKHNLSPAHSINQVPGNS 569
Db 450 FGDRVINLIQATERNIRFRKPP-----IGPIGS-----479
QY 570 GCKFPKDVLMVMEKNGQFAKTAIVAIFILSVASKADAVDAKTKTCTCLLKECRLEAKCI 629

Db 480 -----HVTLVNGNKWASSVEQALGT--LINAFTVYDHHKDSL----- 513
QY 630 SNPACAAVACLOTCNNRPDETECOIKCGDLFENSVDDEFNECAVS--RKCKVPRKSDVGD 688
Db 514 -----TLRGCAANEANYRNKI-----IIVDFSRPLNIPRHVMYPQTEHPFI 554
QY 689 FVPDPDSVLVQKFDKDFSGKWFITRGLNPTF-----DAFDCQLHEFHTEENKLV--- 738
Db 555 FSVISD-----NPTFLNVLVDQSGVERQVLAENYEEGRVAVFG 593
QY 739 ---GNLSWRIRTPDG--GFETRSVQKFCVD--PKYPCILYNHNDNEVLLYODDMVILSSKVE 793
Db 594 KRLSNLK--EYVTLGDYKMFGRFPVQTLPLSRPSRLCASFDQI---KDLIEASKEQ 649
QY 794 NSPEYIFVYVYGRNDAMDGYGSLVYTRSAVLPSIIPELQTAQKVGDRDNTFIKTDN 853
Db 650 NE-----INQMRKRKEAENLELELKVRLKQKRSQAEKVLTTRKELEMHDLK-----N 699
QY 854 TCGPEPPVERL-EKKVEGERTIIKEVEIEEEE---VEKVRD--KEVTL-----FSKLFE 903
Db 700 TVAAE---IESLPSSSVNLOREIMKDLDEIDEKFALEKIQNCLKEAELKANKLTALFE 756
QY 904 GFELQORDE---ENFLRELSEEMDVLDDGLKMEATEVEKL 940
Db 757 NRESAKGEIDAFEAENELKKIEKD-----LQSAEAEKI 791

RESULT 14
Q25662 PRELIMINARY; PRT; 1939 AA.
AC Q25662;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)
DE REPEAT ORGANELLAR PROTEIN.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=96V;
RA Werner E.B., Taylor W.R., Holder A.A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U43145; AAC63403.1; -
SQ SEQUENCE 1939 AA; 229001 MW; B36E462001C6F22F CRC64;

Query Match 2.0%; Score 151; DB 5; Length 1939;
Best Local Similarity 17.5%; Pred. No. 4.1;
Matches 189; Conservative 167; Mismatches 299; Indels 422; Gaps 50;

QY 380 VERLEKTAEGEKKLIKAEVETEEVEKEVEKVRDTEMTLFORLEGPKELOD---BEN 436
Db 95 VREYEVKIELEKEL---KLEKEQINKEYEKELNEKSEFTKQOMELLKEKELNINLEN 151
QY 437 FYRE-----LSKEEKEILNELQME-----ATEVEKLFGRALPIRKLRLMALAPHSNLANHE 487
Db 152 KINNKEITTLKREEK--LNDIESEYIEKNKEKEL---NYEVTNKMSL-----D 196
QY 488 TIKYVGVSKLPDGHKRGKSWGWEDYFGSIVVAKICSSRRIPRYFRKSPRICCGDLDSRLQLF 547
Db 197 KLTCEVQEKKNLEKIN-----KKVIEKENNIRELKEKFWKKEKEILSDG----- 242
QY 548 SHGKHNLSPAHSINQVPGNSGCKPKPDVLMVMEKNGQFAKTAIVAIFILSVASKADA 607
Db 243 -----TIN-----DKKNAYEKLKI-----SPEEKRM 264
QY 608 VDALKTCTCLLKECRLELAKCISNPACAAVACLOTCNNRPDETECOIKCGDLFENSVD 667
Db 265 IEMLD-----SKLIEKEENFAN-----KQAKLE 287
QY 668 EPNCAVSRKKCVPRKSDVGDFFVDPDPSVLVQKFDKDFSGKWFITRGLNPTFADFQCL 727

```

Db 288 KENELIEKLDIESRE-----KDFSK-----EKFASMENEL 321
QY 728 HEFTTENKLVNLS-WRTTPDGGFTTSVAVOKFQVDDPKYFGILYNHNDNEVLLQDDWY 786
Db 322 NTLKSLNACOMEYKLEIKD---LSOSLVKERE-----IFEIKNEY---DD-- 365
QY 787 ILSSKVENSPEDYFVYVYGRNDWDGYSGLYTSRAVLPESIIPELOTAQAKYGRDPN 846
Db 366 ---KINNKEKLSSI---ND--KGIDNTVLHSEEEKI-NKLKKEKETELNEIHKYN 413
QY 847 TFIKTNTCGPEPLVERLEKKVEGERTIIEVEIEEVEKVR-----DKEYT 896
Db 414 LEIE-----TIKNELNEKEEELKKNKAHTVEVNTLTKEIK 449
QY 897 LFSKLEPGKELORDE-----ENFLRELSEEMDVLDGLKWEATEVEKLGRLPIRLKMA 952
Db 450 LLEKKTEDAKEGHKNELNKLNEK-----DNIKNENTEL----- 490
QY 953 VATHCFTSPCHDRIRFFSSDDGIGRLGTRKINGTFLKILPPIQSADLRTGGRSSRP 1012
Db 491 ---NDKISSNSE-----VN-----ILNKDKQILG----- 512
QY 1013 LSAPRSGFGIDIVPLPSKNELKELTAPLLKLVGLACAFIYPSADAVDAKLTAC 1072
Db 513 ---NDIKTLN-----DLINLNK----- 527
QY 1073 LLKGRLELAKIANPACANVACLOTNNRDE-----TECOIKCGDLFENSVD 1123
Db 528 ---EINTSNMKNKEDLAMLNEEMGKC-----VVID 558
QY 1124 EF-----NECAVRKKCVPRK--SDLGEPAPDPVSVLVQNFNSDPNGKWIYITSGLNPTF 1176
Db 559 EIEKKYNEIFMLEELKEKENYADLND---EISILRSIYVKE---KEFI----- 603
QY 1177 DAFDQQLHEPTEGDKNLVGNISWRIKTLDGFTTSVAVOK---FVODNQPGVLNHN 1233
Db 604 ---EMKEFY---ENKI-----NLFNKNFEKKNIYENLNSRLKYDNDQ 642
QY 1234 EYLHYODWYILSSKTIENKPEDYFVYVYGRNDWDGYSGLYTSRAVLPESIIPELEK 1293
Db 643 GLIKQIDELNIQKLEEK---YLQY-----NDN-----MHMFRSCTKIDMYSN 687
QY 1294 AAKSIGRDFST-FI--RTDNTCGPEPALVERIBKTVVEGERIIVKEVEIEEVEKEVEK 1350
Db 688 IKGSDLVDFVTAYIKRDESSDANPDTHK--EMVAELEKRAHAAVLEEKHKEEIAK 745
QY 1351 VGRTEMTLFORLAEGNE---LKQDEENFVRELSEEMFDEIKWEASEVEKLEFG 1403
Db 746 LGEHKEVVLRLGEQHKKEFTIILEEKHKDVYTKLGEQHKENIILKEEHKDVVYTKLG 802

RESULT 15

Q9HAW4 PRELIMINARY; PRT; 1332 AA.
AC Q9HAW4
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HU-CLASPIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21000493; PubMed=11090622;
RA Kumagai A., Dunphy W.G.;
RT "Claspin, a novel protein required for the activation of Chk1 during a
TRL DNA replication checkpoint response in Xenopus egg extracts.";
DR MBL; Cell 6:839-849(2000).
DR EMBL; AF297866; AAG24515.1;
SQ SEQUENCE 1332 AA; 150175 MW; 4EA38DC16A3936C3 CRC64;

Query Match 2.0%; Score 150.5; DB 4; Length 1332;
Best Local Similarity 19.3%; Pred No. 2.6;
Matches 255; Conservative 182; Mismatches 405; Indels 477; Gaps 67;
QY 307 SSQIENKPDYFVYVYGRN-----DAWDGYSGLYTSRAVLPESIIPELPIQKA 355
Db 83 SAEENKEN-----LYAGKNTIKRIYKTVDSDSEYMEKSLYQEN--LEAQVKPCLELS 135
QY 356 AKS-VGRDFNFITDSCGPEPLVERLEKTAEEGKLLIKEAVEIEEVEKEVEKVRD 414
Db 136 LQSGNSTDF-----TDDRKSKKH--IHDKEGTA---GKAVKSKRRLEKE-ERKMEKIRQ 185
QY 415 -----TEMTLFORLEGFKELQ-ODEE-----NFVR 439
Db 186 LKKETKNQEDDVQFPNDSCLLVDKDLFTGLEDDENNSPLEDEESLESRAAVKNKV 245
QY 440 ELSKEKEEILNELQW--EATEVEKLGRLPIRLKMA--LAPHSNFLANHETIKYVGS 495
Db 246 KHKKEPSLESGVHSFEESLSK--GTT---RKERKAARLSKEALKQLHSETQRLRES 300
QY 496 --KLPGHKRFSWGHEDYEGSIVAKICSSRIPRYPRKSPRICGLDSRLG----QLFS 548
Db 301 ALNLPYH-----MPENKTIHDFFRKPRPTCHGNAMALLSKSKYQSS 342
QY 549 HGKHNLSPAHINQVPGNSGCKFPKDVALLMWKWKQFAKTAIVAIFILSVASKADAV 608
Db 343 HHKELIDTANTTEMNSDHHKSGSE-----OTTGAENEVETNA--LPVYSKETQI 389
QY 609 DALKTCTCLLKECLELAKICISNPACANVACLOTNNRDETECOIKCGDLFENSVD 668
Db 390 -----ITGSDSCRDLVK-----NEELEIQEK----- 412
QY 669 FNECAVRKKCVPRKSDVGDPPDPVSVLVQKFDKDFSGKWFITRGLNPTDFADCOLH 728
Db 413 -----QKOSDIRPSP-GDSSVLQOESNF-----LG 436
QY 729 EFHTEENKLVNLSWRIRTPDG-GFFTSRAVQKFEVDP----- 765
Db 437 NNHSECQVGLVAFEPHALEGEQPNPEETDEKVEPEEQONKSSAVGPPKVRRTLDR 496
QY 766 -KYPGILYNHNEVLLYQDDWYILSSKVENSPEDYIFVYVYGRNDWDGYSGLYIY-- 821
Db 497 LKQGVDSIKPRIGADEDSVILEPETNRELEALKORFWKHANPAKPRAGQTVNVNVI 556
QY 822 -----RSAVLPESIIP-----ELQTAQKVG-----RDP----- 845
Db 557 VKDMGTGCKELKADVPVTLAPKLDGASHTKPEGLQVLKAKLQAMKLREERQKR 616
QY 846 NTFIKTDNTCGPEPLVERLEKKVEGERTIIEVEIEEVEKEVEKVRKVTLSKLEFG 905
Db 617 QALFKLDNEDGE-----EEEEEEEMTDESE-EDGEKVEK-EEKEEELEEE--EGK 665
QY 906 KELORDEEN-----FLRELSEEMDVLDGLKWE-----ATEVEKLGRLPIRLKMAVA 954
Db 666 EEEEEEGNQETAEFL--LSSEELTDEKEMKENNDGSGSEIIGKAVGFLSVPKSLSDS 723
QY 955 THCFTSPCHDRIRF-----FSSDDGIGR-----LGITRKRINGTFL--KIL 994
Db 724 TLLLFKDSKSKMGYSPTBEKSETDENSQKPSKLEDDSCSLTKESSHNSFELIGSTI 783
QY 995 PPIQSADLRTGGRSSRPLS-AFRSGSGKIF--DIVPLPSKN--ELKELTAPLLLLKVG 1049
Db 784 PSYQPCNRQTGRGTSFFPTAGGRSP--SPGLFRASLVSSASKSGSKLSEPSLP----- 835
QY 1050 VIACAFILVPSADAVDAKLTACLLKGRLELAKIANPACANVA--CLOTNNRDPETE 1108
Db 836 -----LEDSDLYNASPEPKTLFLG-----AGDFQFCLE-----DDTQ 868
QY 1109 COIKCGDLF-----ENSV---VDEFNECAVSR-----KCVQ 1137
Db 869 SOLLADADGFLVNRHNRNOYQALKPRPLIASMDENAMDMDELDTCTGKFTSOAEKHL 928

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 6, 2001, 02:54:17 ; Search time 6769.47 Seconds
(without alignments)
3630.750 Million cell updates/sec

Title: US-09-075-375A-3
Perfect score: 1589
Sequence: 1 tattttcatgagtttgcaagt.....tctttttcagaagcaaaaaa 1589

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: em_ba1: *
17: em_ba2: *
18: em_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rod: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rod: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

44: em_ov: *
45: em_pat: *
46: em_ph: *
47: em_pl: *
48: em_ro: *
49: em_sts: *
50: em_sy: *
51: em_un: *
52: em_vi: *
53: gb_sts1: *
54: gb_sts2: *
55: gb_sts3: *
56: gb_sy: *
57: gb_un: *
58: gb_vil: *
59: gb_vil2: *
60: gb_htg1: *
61: gb_htg2: *
62: gb_htg3: *
63: gb_htg4: *
64: gb_htg5: *
65: gb_htg6: *
66: gb_htg7: *
67: gb_htg8: *
68: gb_htg9: *
69: gb_htg10: *
70: gb_htg11: *
71: gb_htg12: *
72: gb_htg13: *
73: gb_htg14: *
74: gb_htg15: *
75: gb_htg16: *
76: gb_htg17: *
77: gb_htg18: *
78: gb_htg19: *
79: gb_htg20: *
80: gb_htg21: *
81: gb_htg22: *
82: gb_htg23: *
83: gb_htg24: *
84: gb_htg25: *
85: gb_pr1: *
86: gb_pr2: *
87: gb_pr3: *
88: gb_pr4: *
89: gb_pr5: *
90: gb_pr6: *
91: gb_pr7: *
92: gb_pr8: *
93: gb_pr9: *
94: gb_ro1: *
95: gb_ro2: *
96: gb_in4: *
97: gb_pr10: *
98: em_ba3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1589	100.0	1589	14 NTU34817	U34817 Nicotiana t
2	643.8	40.5	1760	14 LSU31462	U31462 Lactuca sat
3	624.2	39.3	1555	14 AFU44133	U44133 Arabidopsis
4	585.8	36.9	1686	15 SOL250433	AJ250433 Spinacia
5	428.4	27.0	133843	12 AC003981	AC003981 Genomic s
6	426.8	26.9	2884	14 ATU39452	U39452 Arabidopsis
7	419	26.4	1173	13 AF288196	AF288196 Oryza sat
8	162.4	10.2	89479	12 AC006932	AC006932 Genomic s

```

c 9 75.8 4.8 7218 10 I66494
c 10 55.6 3.5 194835 90 AL158158
c 11 54 3.4 181636 69 AC025108
c 12 54 3.4 194615 61 AC010184
c 13 53.4 3.4 178189 64 AC016152
c 14 52.4 3.3 1686 45 E08995
c 15 52.4 3.3 175152 75 AC017819
c 16 52.4 3.3 176157 63 AC013776
c 17 52.4 3.3 182972 68 AC023550
c 18 52 3.3 2254 15 SCYK202W
c 19 52 3.3 3997 10 E14049
c 20 51.6 3.2 165425 71 AC034146
c 21 51.6 3.2 170396 81 AL589703
c 22 51.2 3.2 1141 97 X083744
c 23 50.6 3.2 3121 97 HS088154
c 24 50.6 3.2 3211 97 AR077147
c 25 50.6 3.2 3811 97 HS088153
c 26 50.6 3.2 3901 9 AR077146
c 27 50.4 3.2 4000 14 D83006
c 28 50.4 3.2 177568 92 HS388M5
c 29 50 3.1 152996 87 AC013719
c 30 50 3.1 174989 79 AL355592
c 31 49.8 3.1 75588 90 AL356742
c 32 49.8 3.1 162064 90 AL390882
c 33 49.2 3.1 239130 76 AC079420
c 34 49 3.1 2222 3 SHTLYCA
c 35 49 3.1 5001 96 PFAABRA
c 36 49 3.1 97749 85 AC004906
c 37 48.8 3.1 186738 66 AC021134
c 38 48.6 3.1 5361 9 A58933
c 39 48.6 3.1 5529 96 PFA7010
c 40 48.6 3.1 6152 9 A58932
c 41 48.6 3.1 12240 4 AE001424
c 42 48.6 3.1 175751 62 AC011632
c 43 48.6 3.1 235150 75 AC073776
c 44 48.4 3.0 176372 62 AC012462
c 45 48.4 3.0 188891 69 AC025752

```

ALIGNMENTS

```

RESULT 1
LOCUS NTU34817 1589 bp mRNA PLN 15-JUN-1998
DEFINITION Nicotiana tabacum violaxanthin de-epoxidase precursor (TVDEL),
mRNA, complete cds.
ACCESSION U34817
VERSION U34817.1 GI:1463122
KEYWORDS Common tobacco.
SOURCE Nicotiana tabacum
ORGANISM Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Nicotiana.
REFERENCE 1 (bases 1 to 1589)
AUTHORS Bugos,R.C., Hieber,A.D. and Yamamoto,H.Y.
TITLE Xanthophyll cycle enzymes are members of the lipocalin family, the
first identified from plants
JOURNAL J. Biol. Chem. 273 (25), 15321-15324 (1998)
MEDLINE 98288256
REFERENCE 2 (bases 1 to 1589)
AUTHORS Bugos,R.C. and Yamamoto,H.Y.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-1995) Robert C. Bugos, Plant Molecular
Physiology, University of Hawaii, 3190 Maile Way, Honolulu, HI
96822, USA
FEATURES
source
1. .1589
/organism="Nicotiana tabacum"
/strain="Xanthi"
/db_xref="taxon:4097"
/tissue_type="leaf"

```

```

gene 1. .1589
/genes="TVDEL"
CDS 42. .1478
/genes="TVDEL"
/codon_start=1
/product="violaxanthin de-epoxidase precursor"
/protein_id="AAC50031.1"
/db_xref="GI:1463123"
/translation="MALAPHSNPLANHETIKYVVGSKLPCHKRFSWGWEDYFGSIVVA
KICSSRIPRYFRKSPRICGLRGLQLFSGHKHLNLSPAHSINQVNPKNSGCKFKPK
DVAMWVKWGQFAKTAIVAIFILSVASRADAVDAKLTCTLLKEKRLKLAICSNPA
CAANVACLOQCNNRPDETCIKGDLFNSVDFNECAVSRKKCPVRKSDVDPVP
PDPSELVQKDFMDKDFSGKWFITRGLNPTDFADQLHEFTEENKLVGLNSWRITPD
GGFTFSVOKFVYDPKPYCILYHONEXLYLSDWYILSSKVENSPEDYIFVYKGR
NDAMDGYGGSVLYTRSAVLPESTIPELOTAQKVGDRDFTFKTDNTCCGPEPLVERL
EKVGEGERTIKEVEIEEVEKVRDKVTLFSLFEGFKELQORDEENFLRELSKEE
MDVDGLGKMEATEVEKLFGRALPIRLR"
transit_peptide 42. .443
/genes="TVDEL"
mat_peptide 44. .1475
/genes="TVDEL"
/product="violaxanthin de-epoxidase"
BASE COUNT 500 a 262 c 366 g 461 t
ORIGIN
Query Match 100.0%; Score 1589; DB 14; Length 1589;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1589; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 tattttcatgagtttgagttggtgtaatacaggttggaagaatggtcttgccocatt 60
Db 1 TATTTTCATGAGTTTGCAGTTGGTGGTAATACGTTTGAAGAATGGCTTGTGCCCTCAT 60
Qy 61 caaatcttctgcccacacatgaaacacataattattgttggttcaaatcccggtc 120
Db 61 CAAATTTCTGGCCACACATGAAACCAATCAATATTATGTTGGTCAAGCTTCCCGGTC 120
Qy 121 ataaaggttttagctgggttggaagattacttcttgtagtagtagtagtagtagtag 180
Db 121 ATAAAGGTTTGTAGCTGGGTTGGCAAGATTACTTTGTAGTAGTAGTAGTAGTAGTAG 180
Qy 181 gttccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 240
Db 181 GTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 240
Qy 241 attcaagaggtctctgcaactattctcacacgagggaacacacacacacacacacac 300
Db 241 ATTCAAGAGGTTCTGCAACTATTCTCACACGGGGAACACAACTCTCTCCCGCACATAGCA 300
Qy 301 ttaaccagaattgacccaaaggaattcagagatgcacaaatttccaaagatgtagcttga 360
Db 301 TTAACCAAGAAATGTACCTTAAGGGAATTCAGAGTCAAAATTTCCAAAAGATGTAGCTTTGA 360
Qy 361 tgggttgggagaataatggggcaatttgcacaaacagcaattgtagctatattcatttgt 420
Db 361 TGGTTTGGGAGAAATGGGGCAATTTGCCAAAACAGCAATTTAGCTATATTATTTGT 420
Qy 421 cagttgttccaaagctgagcgggttgatgctctcagagacttgtaacttcttactgaaag 480
Db 421 CAGTTGTCTCAAAAGCTGATGCGGTTGATGCTCTCAAGACTTGTACTTGTCTTACTGAAG 480
Qy 481 agtcaggtttagagcttcgcagagtcatttcgaacctgcagctgcagctgcagctgc 540
Db 481 AGTCAGGTTTAGAGCTTGCAGAGTGCATTTTCGAACCTTGCATGTCAGCTTAATGTTGCT 540
Qy 541 gttccagacttgcaacaatagacctgacacgaaacggaatgtcagataaaatgtgtgtatt 600
Db 541 GTCTCCAGACTTGCACAATAGACCTGACGAAACGGAATGTCAGATAAATGTGTGATT 600
Qy 601 tggttgaaacagtgctcgtagacagtgatcgaatgagtgagtgagtgagtgagtgagtg 660
Db 601 TGGTTGAAACAGTGTCTGTAGACGAGTCAATGATGAGTGTGAGTGTGAGTGTGAGTGTG 660

```

```

QY 661 tacctcgtaaatctgaattgtagtgaactttctgtacctgattccagtgcttctgtccaga 720
Db 661 TACCTCGTAATCTGATGTTGGTGAATCTCTGTACCTGATCCAGTCTTCTTCTCCAGA 720
QY 721 agttgacatgaagattttgacgggaattggttcattactcgcggtttgaatccactt 780
Db 721 AGTTTGACATGAAGATTTTACGCGGAATGGTTCATTACTCGCGGTTTGAATCCCACTT 780
QY 781 ttgatgcttttgattgccaattgcatgagttccatcagagaagaaacaaacttggggga 840
Db 781 TTGATGCTTTTGAATGCCAATGTCATGATGATGATGATGATGATGATGATGATGATGAT 840
QY 841 attatcttgagaaatcagtaacactgagagatgagatgagatgagatgagatgagatgagat 900
Db 841 ATTTATCTTGAGAAATAGTACCTGATGAGGATGATGATGATGATGATGATGATGATGAT 900
QY 901 aattgtgcaagatccaaagtatccggggaactctcaatcatgataatgagatcttc 960
Db 901 AATTGCTGCAAGATCCAAAGTATCCGGGATACCTACATCATGATGATGATGATGATGATGAT 960
QY 961 tctaccaagatgactggtatatttgttoatccaagttagaaaaatagtcagagagattaca 1020
Db 961 TCTACCAAGATGACTGCTGATATTTTGTGTCATCCAAAGTAGAAAAATAGTCCAGAGATTACA 1020
QY 1021 tatttgtactataaaggcagaaatgatgcatgagatgagatgagatgagatgagatgagat 1080
Db 1021 TATTGTGCTACTATTAAGGCGAAGATGATGATGAGGATGATGATGATGATGATGATGATGAT 1080
QY 1081 acacaagaatgctgatttgcctgaaagcattatccgggagttgcaaacccgcagctcaaa 1140
Db 1081 ACACAAGAATGCTGATTTTGCTGAAAGCATTAATACCGGAGTTGCAAAACCGCAGCTCAA 1140
QY 1141 aagttggcggtgatttcaacacattcaataaaaaacagacaatacaatggtgacctgaacctc 1200
Db 1141 AAGTTGGCGGTGATTTCAACACATTTCAATAAAACAGACAATAFACATGTGGCCCTGAACCTC 1200
QY 1201 cccctgttgagaggttgagagaagaagtggagaagagagagagagagagagagagagagag 1260
Db 1201 CCCTTGTTGAGAGGTTTGAAGAGCTCCACAGAGATGAAGAGAACTTCTTAAGAGAGCTGA 1260
QY 1261 ttgaggagatagaagaagaatgagagaggtgagagagagagagagagagagagagagagag 1320
Db 1261 TTGAGGAGATAGAAGAAGAAGTAGAGAAGGTGAGAGATAAAGAACTCACCTTATTCAGTA 1320
QY 1321 aactgtttgagaggttttaaaagagctccaagagatgagagagagagagagagagagagagag 1380
Db 1321 AACTGTTTGAAGGTTTGAAGAGCTCCACAGAGATGAAGAGAACTTCTTAAGAGAGCTGA 1380
QY 1381 gcaagaagaagaatgagatggtttgtgagagagagagagagagagagagagagagagagagag 1440
Db 1381 GCAAGAAGAAGATGAGTGTGATGAGTGTGATGAGTGTGATGAGTGTGATGAGTGTGATGAGT 1440
QY 1441 tttttggcggtgctttaccataaaggaaatttaagtaagttatttttaaaactatcaaat 1500
Db 1441 TTTTGGCGGTGCTTTACCAATAAGGAATTAAGGTAAGTATTTTAAACTATCAACAT 1500
QY 1501 atatacatatgagatgattgtattgtatttctgtgagagagagagagagagagagagagagag 1560
Db 1501 ATATACTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
QY 1561 tgtattgctcttttcagagaagcaaaaaa 1589
Db 1561 TGTATTGCTCTTTTTCAGAGAAGCAAAAAA 1589

RESULT 2
LSU31462
LOCUS Lactuca sativa violaxanthin de-epoxidase (VDE1) precursor, mRNA, 22-JUL-1996
DEFINITION complete cds.
ACCESSION U31462
VERSION U31462.1 GI:1438874

```

KEYWORDS

SOURCE

ORGANISM

Lactuca sativa.
Lactuca sativa
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
II; Asterales; Asteraceae; Lactuceae; Lactuca.

REFERENCE

AUTHORS

TITLE

Bugos,R.C. and Yamamoto,H.Y.
Molecular cloning of violaxanthin de-epoxidase from romaine lettuce
and expression in *Escherichia coli*
Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6320-6325 (1996)

JOURNAL

MEDLINE

AUTHORS

TITLE

Submitted (12-JUL-1995) Plant Molecular Physiology, University of
Hawaii, 3190 Maile Way, Honolulu, HI 96822, USA

JOURNAL

FEATURES

Source

transit_peptide

CDS

mat_peptide

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Score

DB

Length

Indels

Gaps

1;

317

taagggaattcaggatgcaaat

483

TGAGGAATACAAAGATTTGATCTCAAAAGGGCATGACTTTGATTCTTGAAGAAGCAATG

377

gggccaatttgcaaaacagcaattgtagctatattcattttgtcagttgcttcaaaagc

543

GAGACAAATCATCAAAATGGCTATCGTATTTGGTTTGCACATTTGTTATCTTCCAGAGT

437

tgatgcggttgatgctctcaagacttgacttgcttcaagagagagagagagagagagagagag

603

TGATGCGGTGATGCTCTTAAACTTGTCTTGTTCACAAAGATGAGGATGAGCT

497

tgcgaagtgcatttcgaaccctgcagctgtagctaatgttgcctgtctccagacttgcaa

663

TGCAAAATGTATAGCAAAACCCATCTTGTGCGCAAAACGTTGCTGTCTACAGACTTGCAA

557

caatagacctgacgaaacaggaatgtagctataaaatggtggtgattgtttgaaaaacagtg

723

CAATCTCTCTGACGAGACCAATGTGAGATAAAATGTGCTGACTTGTTCGAAAAACAGTGT

617

cgtagacgagttcaatgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag

```

Db 783 GGTGGACCAATTCACAGAGTGTCCGGTTTCCCGAAGAAATGTGTCCCGGAAATTCGGA 842
Qy 677 tgttgtagacttccctgtaactgataccagtgctctgttccagaaagtttgacatgaaaga 736
Db 843 TGTGGGTGAATTCCTGGTTCGGATCGTAATACAGTGGTGTGAAATTTTAAACATGAAGA 902
Qy 737 tttagcgggaaatgttcaactactcgcggtttgaatccacacttttgatgcttttgatg 796
Db 903 CTTTAGTGGGAAGTGGTATATACAAAGTGGTTTAAATCTCTACATTTGATGCAATTTGAT 962
Qy 797 ccaattgcatgagttccatcagaaagaaacaaactgttggggaattttatcttggagaat 856
Db 963 TCAACTTCATGAGTTCATATGAAAATGATAAATCTGTGGGAACCTTAACATGGCGCAT 1022
Qy 857 acgtacacatgtagggagatttttactgagatcagcgggtgcacaaatctgccaagatcc 916
Db 1023 AAAAAGTTTGGATGGTGGTTCCTTTTACTCGATCTGTGCAAAACATTTGTTCAAGATCC 1082
Qy 917 aaagtatccgggataactctacatcatgataatgagtagtctctctcacaagatgactg 976
Db 1083 AGATCTTCCTGGAGCATTATATCATGACATGAGTTCTTCACATACCAAGATGACTG 1142
Qy 977 gtatatttgcatacgaagtagaataatagtcacagaggattacatatattgtgtactataa 1036
Db 1143 GTACATATATCTTCCCAATCGAAAACAAACCCGATGATTACATATTGCTATACTACCG 1202
Qy 1037 ggcagaaatgacatggatgagatgattggtgtgtctgtactttacacaaagtgactg 1096
Db 1203 AGTTCGAAACGACGATGGGATGATACGTCGGTCCGATCTACACCCGAAGCCCGAC 1262
Qy 1097 ttgcttgaagcattataccgagtttgcacacgcagctcacaagagttggcggtgattt 1156
Db 1263 ACTCCCGAATCGATCATCTCCAAACCTTACAAAAGCAGCCAAATCCGTGGTGCAGACTT 1322
Qy 1157 caacacattcataaaacagacatacatatggtccctgaacctccctctgttgagaggtt 1216
Db 1323 TAACAATTTTCAACAAACCCGACAAATAGTTGTGGGCGCTGAGCCTTCCATTTGGTGAAAGGCT 1382
Qy 1217 ggaagaagaagtgaagaaggaagagcagatcataaagaagttgagagagagaaga 1276
Db 1383 TGAGAAACAGCGGAGAGCGGAGAGTGTGATATGAAGAGAGCTGTAGAGATAGAAGA 1442
Qy 1277 agaagta-----gagaagtgagagataaagaagtcacattatcagtaaaact 1324
Db 1443 AGAGGTTTGAAGAGAGGTGGAGAGGTGTAGAGATCTGAGATGACTTTGTTTCAGAGGTT 1502
Qy 1325 gttgaaaggttttaaagagctccaagcagatgaagagaactctttaaagagagctgagcaa 1384
Db 1503 GCTTGAAGGTTTAAAGAGTGTGCAACAGATGAAGAGAATTTTGTGAGGGAGTTGAGTAA 1562
Qy 1385 agaagaatggatgttttggatggacttaaaatggaagcaactgaggtgagaacactttt 1444
Db 1563 AGAAGAGAAGGAAATTTCTGAATGAACCTTCAATGGAAGCGACTGAAGTTGAAAGGCTTTT 1622
Qy 1445 tggcgctgtttaccataaagaaataaagtagtaagttatttttaaaactatcaacatatat 1504
Db 1623 TGGGCGCGGTTACCGGATAGGAAGAACTTAGATAAAATTTTCGATGATTGATTCAGACAATAT 1682
Qy 1505 a 1505
Db 1683 A 1683

```

RESULT 3

*ATU44133

LOCUS

DEFINITION

Arabidopsis thaliana violaxanthin de-epoxidase precursor (AVDE1)

mRNA, complete cds.

ACCESSION

U44133

*VERSION

U44133.1

GI:1465734

KEYWORDS

thale cress.

```

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 1555)
AUTHORS Bugos,R.C., Hieber,A.D. and Yamamoto,H.Y.
TITLE Xanthophyll cycle enzymes are members of the lipocalin family, the
first identified from plants
J. Biol. Chem. 273 (25), 15321-15324 (1998)
JOURNAL 98288256
MEDLINE 2 (bases 1 to 1555)
REFERENCE Bugos,R.C. and Yamamoto,H.Y.
AUTHORS Direct Submission
TITLE Submitted (02-JAN-1996) Robert C. Bugos, Plant Molecular
JOURNAL Physiology, University of Hawaii, 3190 Maile Way, Honolulu, HI
96822, USA
FEATURES
source Location/Qualifiers
1..1555
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
1..1555
/gene="AVDE1"
45..1433
/gene="AVDE1"
/codon_start=1
/product="violaxanthin de-epoxidase precursor"
/protein_id="AAC50032.1"
/db_xref="GI:1465735"
/translation="MAVATHCFTSPCHDRIRFFSSDDGIGRLGTRKINGTEFLKIL
PPQSADLTGGRSSPLAFSGFKSLGDIVLPFSKNELKELTAPLLKLVGVLA
CALIVPSADAVDALTKACLLKGRLEAKCIANPACANVACLQTCNNRPDETBCQ
IKGDLFENSVDDEFNECAVRKCPKSDLGEPAPDPSVLVQNFNISDFNGKWI
TSGLNPTFDADQLHEFTEGDKNLVGNISWRKTLDSGFFTRSAVQKVPQDPNPG
VLNHDNEYLHYODDWVILSSKIENKPEDYFVYGRNDAMDGYGAVVYTRSSVLP
NSIPELEKAAKSGIDFSTFIRTDNTCGPEPALVRIEKTVEEGERIIVKEVELEE
EVEKEVEKVGRTMTLQRLAEFGNELKQDEENFVRELSEKEMEFLDEIKMEASEVEK
LFGKALPIRKVR"
transit_peptide 45..383
mat_peptide /gene="AVDE1"
384..1430
/gene="AVDE1"
/product="violaxanthin de-epoxidase"
BASE COUNT 485 a 285 c 376 g 409 t
ORIGIN
Query Match 39.3%; Score 624.2; DB 14; Length 1555;
Best Local Similarity 74.3%; Pred. No. 1.3e-123;
Matches 820; Conservative 0; Mismatches 268; Indels 15; Gaps 2;
Qy 424 ttgtctcaaaagctgtagcgggttgatgctctcagaacttgacttgcttactgaaagagt 483
Db 364 TTGTTCCCATCTGCAGATCAGTTGATGCACCTTAAACACTTGTGCATGCTTATTTGAAGGAT 423
Qy 484 acagattagactgtagcgaagtgacatttcgaacccctgcatgtgcagctaaattgttcctgtc 543
Db 424 CGAGGATAGAAGTCCCAAGTGCATTCGCAACCCCTTCCTCTGCGACCCCAAGTCCCGTGC 483
Qy 544 tcagacttgcaacaatagacactgacgaacggaatgtcagataaaatgtgtgattgtt 603
Db 484 TTCAGACCTGCAATTAACCGTCCAGATGAACCGAGTGCAGATTAAATGTGGGATCTGT 543
Qy 604 ttgaaaaacagtgtagcagagttcaatgagtgtagtgcagctcccgaaagaatgtgtac 663
Db 544 TTGAGAAACAGTGTGTTGATGATTCACAGAGTGTGCTGTGTCGAGAAAAAAGTGTGTTC 603
Qy 664 ctctgaaatctgattgttggtgacttctctgactgaacccagtggtcttctgtccagaagt 723
Db 604 CTAGAAAATCTGATTCGAGAAATTTCTCTGCCCGACACCCCTTCCTGTTCTTGTACAGAACT 663
Qy 724 ttgacatgaaagatttttagcgggaaatggttcattactcgcggtttgaatccccactttt 783

```


QY 881 tactcatcagcgggtcgaataattcgtgcaagatccaaagtatccgggagatctactacaa 940
 Db 810 CACACGACTGCTGTACAGAAATTTCCGCAAGACCCCTCAACCTGGGAATCTCTATAA 869
 QY 941 tcagtataatgagtgatctctctaccagatgactgggtatatttggatccaaagtaga 1000
 Db 870 TCATGACATGCATATCTTCACATCAAGATGATGGTACATCCTATCTTCAAAATTGA 929
 QY 1001 aaatagtcagagagattacatatcttggctactataaggcagagaatgatcgatggatgg 1060
 Db 930 AAATCAACAGATGACTAGTATTTGTATATTACCAGCAGGAATGATCGTGGGATGG 989
 QY 1061 atagtgtgtctactttacacagaagtcagtttggctgaaagcattataccaga 1120
 Db 990 CTATGTGGGGCATTTCTTTACACAAGAGTGCAACTGTACCTGAAATATCTGCTCTGA 1049
 QY 1121 gttgcaaacgcagctcaaaaagttggcgtgatttcaacacattcatataaacagacaa 1180
 Db 1050 ACTTAACAGAGCAGCTCAAAAGTGTAGGAAAAGACTTCAATAATTCATCCGAACGACAA 1109
 QY 1181 tacatgtgcccgtgaacctccctcttggagaggttggagaagaagtgaagaaggaga 1240
 Db 1110 TACTGTGGGCCCGACCTCCACTGTTGAGAGCTGGAGAGACGGTGGAGGAGGAGA 1169
 QY 1241 aaggagcatcaaaaagaagtga-----ggagatagaagaagaagttagagaa 1288
 Db 1170 AAGAACTATCATTAAGGAAGTAGAACAAATTAGAAGGAGAGATAGAGGAGACCTGGAAA 1229
 QY 1289 ggtgagagataaagaagtcaccccttctcagtaaacctgttgaagttttaaagagctcca 1348
 Db 1230 GGTGGGAAAACCTGAGATGACATTTGTTTCAGAGCTTACTAGAGGTTTTCAGAGCTTCA 1289
 QY 1349 acagatgaagaagaactcttaagacagctgagcaagaagaagaagtgttttgatgg 1408
 Db 1290 AAGAGATGAAGATATCTCTCTCAAGAAATTAACAAGGAGAGCGGAGAAATTACTGGAAGA 1349
 QY 1409 acttaaatggaagcaactgagtagaagaactttttggcgtgctttaccataaggaa 1468
 Db 1350 CTAAAGATGCGACCTGGCGAGCTGGGAAAACCTATTGTTGGCTAGCATTTACCTATAAGAAA 1409
 QY 1469 attaaagtaagtatt 1483
 Db 1410 GCTTCGGTAACATTT 1424

RESULT 5
 AC003981
 LOCUS AC003981 133843 bp DNA PLN 05-JAN-2001
 DEFINITION Genomic sequence for Arabidopsis thaliana BAC F22013 from chromosome I, complete sequence.
 ACCESSION AC003981
 VERSION AC003981.2 GI:9454484
 KEYWORDS HTG.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 133843)
 AUTHORS Khan, S., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C., Kim, C., Shinn, P., Altafi, H., Bei, Q., Chin, C., Chlou, J., Choi, E., Conn, L., Conway, A., Gonzales, A., Hansen, N., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Federspiel, N., Theologis, A., and Ecker, J.R.
 TITLE Genomic sequence for Arabidopsis thaliana BAC F22013 from chromosome I
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 133843)
 AUTHORS Ecker, J.R.
 TITLE Direct Submission

JOURNAL Submitted (08-JAN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
 REFERENCE 3 (bases 1 to 133843)
 AUTHORS Ecker, J.R.
 TITLE Direct Submission
 JOURNAL Submitted (26-JAN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
 REFERENCE 4 (bases 1 to 133843)
 AUTHORS Ecker, J.R.
 TITLE Direct Submission
 JOURNAL Submitted (18-APR-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
 REFERENCE 5 (bases 1 to 133843)
 AUTHORS Cheuk, R., Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C., Khan, S., Kim, C., Altafi, H., Bei, B., Chin, C., Chlou, J., Choi, E., Conn, L., Conway, A., Gonzales, A., Hansen, N., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R., Federspiel, N., Theologis, A., and Ecker, J.R.
 TITLE Direct Submission
 JOURNAL Submitted (18-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
 REFERENCE 6 (bases 1 to 133843)
 AUTHORS Ecker, J.R.
 TITLE Direct Submission
 JOURNAL Submitted (26-JUL-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
 REFERENCE 7 (bases 1 to 133843)
 AUTHORS Ecker, J.R.
 TITLE Direct Submission
 JOURNAL Submitted (29-JUL-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
 REFERENCE 8 (bases 1 to 133843)
 AUTHORS Cheuk, R., Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C., Khan, S., Kim, C., Altafi, H., Bei, B., Chin, C., Chlou, J., Choi, E., Conn, L., Conway, A., Gonzales, A., Hansen, N., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R., Federspiel, N., Theologis, A., and Ecker, J.R.
 TITLE Direct Submission
 JOURNAL Submitted (15-AUG-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
 REFERENCE 9 (bases 1 to 133843)
 AUTHORS Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C., Khan, S., Kim, C., Altafi, H., Bei, B., Chin, C., Chlou, J., Choi, E., Conn, L., Conway, A., Gonzales, A., Hansen, N., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R., Federspiel, N., Theologis, A., and Ecker, J.R.
 TITLE Direct Submission
 JOURNAL Submitted (05-JAN-2001) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
 COMMENT On Jul 26, 2000 this sequence version replaced gi:3063438.
 FEATURES
 source
 1. 133843
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="1"
 /clone="F22013"
 join(189, 319, 505, 605, 731, 876)
 /note="hypothetical protein"
 /codon_start=1

CDS


```

/evidence-not_experimental
/product="F22013.1"
/protein_id="AAF99751.1"
/db_xref="GI:9802549"
/translation="MGRATVFAKLADTARKELPSTLAAVRVSGMEISDLTULESLDS
QDITGINKAKVAQAAEAGIKQIGTLAQOQTLSMIEERANLPEISLQPVVAGAETK
SHAIGATKRLMNIITGGNDED"
join(1385..1463,2004..2185,2419..3015,3461..3691,
3803..3943,4030..4230,4333..4620)
/notes="similar to sigma factor 2 gb|AAB69385.1; similar
to EST dbj|AV557315.1"
/codon_start=1
/evidence-not_experimental
/product="F22013.2"
/protein_id="AAF99752.1"
/db_xref="GI:9802550"
/translation="MSSCLLPQFKPPDSFHSIFRTSFCAPKHNKGSVFFQPCAVST
SPALLTSLMDVAKLRFSDFTSDSLSDISDQWYTRPDGPFSTEAKYLEALASETLTS
DEAVVAAAAGAAVALAAVAKADATLTKNSNTNLLTSSADRSKWDQFTEKERA
GILGLAVSDNGIVSDKITASANKESIGDSEKOEVELLEEOPSVSLAVRSTROT
ERKARAKGLEKTASGIPSVKTSKPKRLVAQEVHDNDPLRLMTSSSKLLTVR
EHELASGIDQLKLELQELTSGRQPTFAWASAAGVDOKSLURKHGHTCKD
KMIKSNRLVISIAKYNQAGMNLQDLVQEGCGRLVGAKEFDATGKFESTYAHMWI
KQAVRSLDSQSMIRLPHFMVATYRVKEARQLYSETGKHPKNEIEAATGLSMKR
LMAVLLSPKPPRLSDQKIGNONLKPSEVADPEAVTSEDILIKEMRODLKVLDSL
GTREKQVIRWRFMEDGRMKTLOEIGEMGVSRERVQRIESSAFRLKKNKRNHLOQ
YLVAAQS"

```

CDS

CDS

```

SMCYSLGVHVLPPFFRYRGSGRVCFSCTNATIKKFRDALAKHGPDRCSLGPTKGLE
EKELVALAANKELNFTYTPKVPVKEKATPDSPSLPPLPSMSSNDEKTLVSGAR"
complement(join(13219..13280,13397..13477,13866..13932,
14113..14274))
/notes="unknown protein; similar to EST dbj|AV530752.1"
/codon_start=1
/evidence-not_experimental
/product="F22013.6"
/protein_id="AAF99782.1"
/db_xref="GI:9802580"
/translation="MAKNEIAKKREKAKRRAIHGDPDPLTNKLTPTPVSVSGKR
QRKLLKWRREKQKMGKLVMTMEDVEMASQAASEDSKSPKFSVKSLKLKLNIN
KGGKKKNOKASGEKSGADCMLE"
16782..19871
/notes="similar to receptor protein kinase-like protein
emb|CAB66905.1; similar to ESTs dbj|AV522839.1,
dbj|AV528375.1, and dbj|AV541093.1"
/codon_start=1
/evidence-not_experimental
/product="F22013.7"
/protein_id="AAF99755.1"
/db_xref="GI:9802553"
/translation="MAIPRLFLFYIIGFALFPFVSSEFQNSBOEILLAKPSDLFDP
SNLQDWKRPENATTFSELVCHTGWCHDANGYVAKLLSNMNLGNSVQIQSFPS
LOALDSNAPPESSLPKLSLTKVIDSVNSPFGTFPYGLMATGLTHVNASSNN
ESGFLPEDLGNATLLELDVDFRGYFEGSVPSFKNLKFLGLSGNNFGKVPKVIQ
EUSLETTIILYNGFMGEIPEFGKLRLQVLDLAVGNLTQIPISSQLQKLTIVYL
YONRLTGKLPRELGMSTSLVLDLSDNQITGEIPEMVGELKNLQNLNMRNLQITLIP
SKIAELPLEVLELWQNSLMGSLPVLHKNPSPLKWLVDYSSNKLSDIPSGLCYSRNL

```

CDS

```

join(5386..5601,5679..5845,6111..6207,6312..6706,
6798..7311)
/notes="similar to viiolaxanthin de-epoxidase
emb|CAB59211.1; similar to ESTs gb|A1994713.1,
dbj|AV440942.1, and dbj|AV521985.1"
/codon_start=1
/evidence-not_experimental
/product="F22013.3"
/protein_id="AAF99753.1"
/db_xref="GI:9802551"
/translation="MAVATHCFTSPCHDRIRFSSDDGICRLGITRKNNGFLKIL
PPTQSLDRITGSRSRPLSAFRSGSKGIFLDIVPLPSKNELKELTAPLLKLVGVIA
CAELVPSADVALTACCLKGRKRIELAKIANPACANVACLOTQNNRPDETECO
IKCGDLPFSAVDFNECAVSRKKVPSKSLGELFPADPFSVLQNFNIDFNKWIY
TSGNLTFFADQCLHEFHFPKNSDLGKLVGNISWKIKTLDGFFTRSAVQKFOVDNPGI
VLNHDNELYHODDYILSSKIKENPEDYIFVYVYGRNDAMDGYGSAVYVTFSSVLP
NSIPELEKAAKSGIDRDSFPIRTDNTCGPEPALVERIEKTEVEEGRILVKEVELEE
EVEKEVKGRTMETLQRLAEGFNELKQDEENFVRLKEEMEFLEIDINMEASEVEK
LFGKALPIRKVR"

```

CDS

```

Query Match      27.0%; Score 428.4; DB 12; Length 133843;
Best Local Similarity 67.6%; Pred. No. 1.3e-81;
Matches 703; Conservative 0; Mismatches 231; Indels 106; Gaps 3;

QY 578 atgcagataaaatgtggtgattgtttgaaaacagtgctgtagacgagttccaatgagt 637
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6305 ATATCAGATTAAATGTGGGAGTCTGTTTGTGAGAACAGTGTGTTGTGAGTGTCAACGAGTG 6364

QY 638 tgcagttctcccaaaaataatgtacctcgtataatctgattgtgtgaccttctctgtaac 697
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6365 TGCTGTGTGCGAGAAAAAGTGTCTCTAGAAAATCTGATCTCGGAGAAATTTCTCTGCCCC 6424

QY 698 tgcaccagdtgtcttctgcacaaagtctcacatgaaagtatttagcggaaatggttcat 757
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6425 AGACCCTCTGTTCTTGTGACAGAACTTCAACATCTCGGACTTTAAACGGGAAGTGTGATAT 6484

QY 758 tactcgcggtttgaatcccaacttttgcattgttgcattgcattgcattgcattgcattgc 817
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6485 TACAAGTGGCTTGAATCCAACTTTTGTGCTTCGACTGCCAGCTGCATGAGTTCACACAC 6544

QY 818 agaa---gaaacacaaatgtgggggaatttatcttggagaatacgtacacctgattgagg 874
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6545 AGAAGGTGACACAAAGCTTGTGGAAACATCTCTTGGAGAAATAAAGACCTTACACAGTGG 6604

QY 875 atttttactcgaatgacgagtgcaaaaattcgtgcaagaatcccaagaatcccgaggatct 934
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6605 ATTCTTTTACTAGTAGCCCGTACAAAATTTTCGGAAGATCTTAAACCAACCTGGTGTCT 6664

QY 935 ctacaatcatgaatgagtgatctctctaccagaatgact----- 975
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6665 CTACATCATGACACAGGATACCTTCACATCAAGATGACTGGTAATAACATCTTGTATC 6724

QY 976 ----- 975
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6725 CACCCACAGGAGGAGTCTTTGTATGATGCTGCTATTTCTTTTGGCCTAAACAGCTTTTCAT 6784

QY 976 -----gggtattttgtcatccaaagtaaaaaatagtcacagaggattacatat 1023
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6785 ATGATTTTTCAGGTATATCTCTCATCAAGATGAGTAAACCTGAAGACTATATAT 6844

QY 1024 ttgtgtactataagggcagaataatgcatgagtgatgattgattgattgattgattgattgatt 1083
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6845 TTGTATACCTCCGCGGCGAAGCATGCTTGGGATGATATGTTGCGAGTTGTATACA 6904

```

CDS

```

complement(join(7702..7832,7895..8606))
/notes="similar to s-syntaxin-like protein
gb|AA00648.1|AC009540.25; similar to ESTs
dbj|AV556140.1, dbj|AV566414.1, and dbj|AV555627.1"
/codon_start=1
/evidence-not_experimental
/product="F22013.4"
/protein_id="AAF99783.1"
/db_xref="GI:9802581"
/translation="MSYVDLKKAAKMDMEAGPFDLEMASTRADKMDENLSFLEPAE
YVKAEMGLISETLAEYHEESKGVHKAESVRLRNKINSNEIVSGLRKAKSTKSLKE
EMDKANKEIKRLSGTPYRSRTAVTNGLRKLKEVMEFQGLRQKMSYKETVERRY
FTVTKHEANDMEIKETITNAGGEELFTRAIQHGKGVKLTVEIQDRYDAKAELEK
SLELHQVFLDMVAVESQANELKTAKSHQRNKKWMCIGIIVLLIILVIVPIITS
FSSS"

```

CDS

```

join(10982..11218,11470..11646,11732..11896,11987..12235)
/notes="similar to thioredoxin-like 1
gb|AAD35005.1|AF144387.1; similar to ESTs gb|T46281.1,
dbj|AV442467.1, and dbj|AV549179.1"
/codon_start=1
/evidence-not_experimental
/product="F22013.5"
/protein_id="AAF99754.1"
/db_xref="GI:9802552"
/translation="MTEVISTSLIFLACGNHHRVDDFSFSPVSGFGGLIKSFSCLK
LKSQELPVRVYKQIVGSDQSEFSRRAITAOITLRTIGTAKWKEGLKDNMREI
SSAQELVDSLTNAGDKLVVYDFFSPGCGGCKALHPKICQPAENNPVQVLOVNYEBKH

```

QY 1084 caagaagtgcagtttgcctgaagcattatccggagttgcaaacgcagctcaaaaag 1143
 Db 6905 CGAAGAGTCTGTATTACCAATAGCATTATACAGAACTCGAAAAGCAGCAAAAAGCA 6964
 QY 1144 ttggcgctgatttcaacacattcataaaacagacatacatgtggccctgaacctcccc 1203
 Db 6965 TAGCAGAGACTTTCAGCACATTCATTAGAAGCGATTAACACATGTGGTCTGAACTGGCG 7024
 QY 1204 ttgtgaagttggaagaagaagtgaagaagaagaagaagaagaagaagaagtgg 1263
 Db 7025 TCGTGGAGAGATTGGAAGACAGTGGGAAGAGGTGAAGAGATAATCGTAAAGAGGTTG 7084
 QY 1264 agggatagaagaagaagtgaagaagtgaagataa-----agaagtcacct 1311
 Db 7085 AAGAGATAGAAGAAGAGTGAAGAAGAGTGGGAAGGTCGGTAGGACTGAGATGACCT 7144
 QY 1312 tattcaagtaactgttgaagggttttaaaagagctccaaagagatgaagaagacttttaa 1371
 Db 7145 TGTTCCAGAGATTGGCTGAAGGATTTAATGAAGTGAAGCAAGAGAGGAAATTCGTGA 7204
 QY 1372 gagagctgagcaagaagaagtgtgttttggatggacttaaaatggaagcaactgagg 1431
 Db 7205 GAGAGTTAAGTAAGAAGAGATGAGGTTTTTGGATGAGATCAAAATGGAAGCAAGTGAGG 7264
 QY 1432 tagaaaaacttttggcgctgtttaccataaggaataaaggttaaggtattttttaa 1491
 Db 7265 TTGAAAAATTTGTTGGGAAAGCTTTGCCAATCAGGAAGGTTCAGGTGAACAAAGAACCCAC 7324
 QY 1492 tatcaacatatatactacat 1511
 Db 7325 CATTTGTTGTAACAACTATAT 7344

RESULT 6
 ATU39452/c
 LOCUS Arabidopsis thaliana syntaxin-related knolle mRNA, complete cds.
 DEFINITION U39452
 ACCESSION U39452.1 GI:1184166
 VERSION
 KEYWORDS
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 2596)
 Lukowitz,W., Mayer,U. and Jurgens,G.
 Cytokinesis in the Arabidopsis embryo involves the syntaxin-related
 KNOLLE gene product
 Cell 84 (1), 61-71 (1996)
 JOURNAL 96140640
 MEDLINE
 REFERENCE 2 (bases 1 to 2884)
 AUTHORS Lukowitz,W.
 TITLE Direct Submission
 JOURNAL Submitted (26-OCT-1995) Wolfgang Lukowitz, Lehrstuhl fuer
 Entwicklungs-genetik, Universitaet Tuebingen, Spemannstrasse 37-39,
 D-72076 Tuebingen, Germany

FEATURES
 source Location/Qualifiers
 1..2884
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 520..1452
 /gene="knolle"
 520..1452
 /note="syntaxin-related"
 /codon_start=1
 /protein_id="AAC49163.1"
 /db_xref="GI:1184167"
 /translation="MNDLMTKFSYVDLKKAAKMDMEAGPDFDLEMASTKADKMDEN
 LSSFLFEAEYVKAEMGLISLTARIEQYHEESKGVHKAESVKLRNKSINETSGLRK

AKSKSLEEMDRANKERIKRISGTPVYRSRTAVTNGRLKRLKEVMEFOGLRQMKSE
 YKEVERRYFTVTGEHANDMEIEKIITDNAGKEEFLTRAIOEHGKGVLTVEIODR
 YDAKETEKSLLELHQVFLDMAYVESQGMDEIEHHVINASHYVADGANELKTAKS
 HORSNRKMCIGITVILLIILIVVPIITSSSS"

BASE COUNT 833 a 569 c 582 g 900 t
 ORIGIN

Query Match 26.9%; Score 426.8; DB 14; Length 2884;
 Best Local Similarity 67.5%; Pred. No. 2.3e-81;
 Matches 702; Conservative 0; Mismatches 232; Indels 106; Gaps 3;

QY 578 atgcagataaaatgtggtgattgtttgaaaacagtgctgtagacgaggttcaaatgagt 637
 Db 2849 ATATCAGATTAAATGTGGGATCTGTTTGAGAACAGTGTCTTGTGATGAGTCAACGAGTG 2790
 QY 638 tgcagtcctccgagaagaatgttacctcgttaaatctgataatgttggtagacttctgtacc 697
 Db 2789 TGCTGTGTCGAGCAAAAAGTGTGTTCTAGAAAATCTGATCTCGGAGAAATTTCTGTGCC 2730
 QY 698 tcatcccgagttctgtccgagaagtgttgacatgaaagatttttagcggaagtgttcat 757
 Db 2729 AGACCCCTCTGTTCTGTACAGAACTTCAACATCTCGGACTTTAACGGGAAGTGTACAT 2670
 QY 758 tactcgcggtttgaatcccaacttttgatgcttttgattgccaattgcatgaggttccatac 817
 Db 2669 TACAAGTGGCTTGAATCCAACTTTGTATGCTTCGACTGCCAGCTGCATGATGAGTTCACAC 2610
 QY 818 aaaa---gaaaacaaactgtggggaatttatcttgagaatacgtacacactgagagg 874
 Db 2609 AGAAGGTGACACACAGCTTGTGGNAACATCTCTTGGAGATAAAGACCCCTAGACAGTGG 2550
 QY 875 atttttactcgatcgcggtgcaaaaattcgtgcaagatccaaagtatccggggatct 934
 Db 2549 ATTCTTTACTAGTCAGCCGTACAAAAATTCGTGCAAGATCCTAACCAACCTGGTGTCT 2490
 QY 935 ctacaatcatataatgattcttcttccaaagatgact----- 975
 Db 2489 CTACAATCATGACACGAGTACTCTTCACTCAAGATGACTGGTAATAACATCTTGATC 2430
 QY 976 ----- 975
 Db 2429 CACCACGAGGAGAGTCTTGATAGTGTGCTATTCTTTTGGCTTAACAAGCTTTTCAT 2370
 QY 976 -----ggatatattgttcataccaaagtagaataatagtcagagagattacatat 1023
 Db 2369 ATGATTTTTCAGGTTATCTCTGTCATCAAGATAGAGAATAAACCCTGAAGACTATATAT 2310
 QY 1024 ttgtgtactataaaggcgcaagaatgatgcatgggtggatggataggttctgtactttaca 1083
 Db 2309 TTGTATACTACCGTGGCGCAACGATGCTTGGATGGATATGGTGGTGCATTTGTATACA 2250
 QY 1084 caagaagtgcagtttgcctgaagcattatccggagttgcaaacgcagctcaaaaag 1143
 Db 2249 CGAGAAGTCTGTATTACCAATAGCATTATACAGAACTCGAAAAAGCAGCAAAAAGCA 2190
 QY 1144 ttggcgctgatttcaacacattcataaaacagacatacatgtggccctgaacctcccc 1203
 Db 2189 TAGGCAGAGACTTCAGCACATTCATTAGAACGGATACACATGTGGTCTGCTGAACTCGGC 2130
 QY 1204 ttgtgagaggttgagagaagaagtgaagaagagagagagagagagagagagagagag 1263
 Db 2129 TCGTGGAGAGATTGAGAAGACAGTGGGAAGAGGTGGAAGGATAATCGTAAAGAGGTG 2070
 QY 1264 agggatagaagaagaagtgaagaaggtgagagataa-----agaagtcacct 1311
 Db 2069 AAGAGATAGAAGAAGAGGTAGAGAAGGAAAGTGGAGAGAGGTGGTAGGACTGAGATGACCT 2010
 QY 1312 tattcaagtaactgttgaagggttttaaaagagctccaaagagatgaagaagacttttaa 1371
 Db 2009 TCTTCCAGAGATTGGCTGAAGGATTTAATGAAGTGAAGCAAGACGAGAGAGAAATTCGTGA 1950
 QY 1372 gagagctgagcaagaagaagtgtgttttggatggacttaaaatggaagcaactgagg 1431

/note="putative ribosomal protein L10 gi|3860277: similar to ESTs gb|T41543, gb|R86781, gb|AA721952, emb|Z34253, gb|N64989, gb|AA041179, emb|Z26567"
/codon_start=1
/evidence=not_experimental
/product="T27G7.6"
/protein_id="AAF22886.1"
/db_xref="GI:6664304"
/translation="MSKQSEAVREAITITGKSEAKRNFVETIELQIGLKNYDPQK DKRPSGVKLPIPRPMKICMIGLDQAHVEAEKMGLENMDVESLKKLNKKLVKLL AKYHAFVASEYIKQIPRLGLNKGAKGFWLQILFHFCNLSMEEEKQIFQNVQMSVN FLVSLKKNQNVRCILYKSTMGPPQRI"
join(19189, 19347, 19428, .19605, 19741, .19885, 20087, .20143, 20295, 20409, 20600, 21079, 21145, 21201)
/note="similar to yeast dcpi gi|3169719: similar to ESTs gb|H37036, emb|Z27040, gb|R90173, gb|A1999440.1"
/codon_start=1
/evidence=not_experimental
/product="T27G7.7"
/protein_id="AAF22887.1"
/db_xref="GI:6664305"
/translation="MSQNGKILPNLDONSTRLLNLVLRIDPYIEELITAAHVTFY EFNIELSQSRKDVESLGVYKRLLLSIYNYVLIPNRLKSDWILFFVFSFSTQPRF QFTVMNRNTDNVLLGLDFEYVQGPYLLYNASQEVNGIWFYNKRCEVEVATLFN RILSAYSKVNOKPASKSEFELEAKPTMAVMDGLPSPSTARDPDPAFVNFPS STNLGNATSGSAGPYQSSAIPHOPHOPTIAPPVAAAAPPQIQSPDQLQSSPL MTLFDNPNVISNSNIHTDLYTPSFSGPPRMAQPHLIPGVMSPTAPEPLNNASHQ QRSYGVPLVQFPFPPLPPSLAPATPGPVISDKVKEALLSLQEDFIDKTRITLON ALQO"
complement(join(24182, .26122, 26212, .26300, 26373, .26501, 26545, .26839))
/note="similar to Inerleukin-1 beta convertase homolog A precursor sp|P55865"
/codon_start=1
/evidence=not_experimental
/product="T27G7.8"
/protein_id="AAF22903.1"
/db_xref="GI:6664321"
/translation="MLPMEHPLPLSFVLPNPLTGLALGFDIGNFVDLQLLLRASLT SNLNHDCSDLNDRLLHURDITKHAVSWISTLSAKVSLIEDRLNLSLCLHIEF

Query Match 10.2; Score 162.4; DB 12; Length 89479;
Best Local Similarity 77.8; Pred. No. 1.le-24;
Matches 196; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 578 atgcagataaatgtggtgattgtttgaaacagtgctgtagacaggttcaatgagt 637
Db 89221 ATATCAGATTAAATGTGGGATCTGTTTGAGACAGTGTGTTGATGAGTCAACGAGT 89280
QY 638 tgcagctcccgaaagaaatgtcacctgtaaatctgattgttgtaacttctctgacc 697
Db 89281 TGCTGTGTCGAGAAAAGTGTCTCTAGAAAATCTGATCTCGGAGAAATTTCTGCCCC 89340
QY 698 tgatccagtggtcttctccagagatttgacatgaaagatttagcggaatggttcac 757
Db 89341 AGACCCCTCTCTTCTTACAGAACTTCAACATCTCGGACTTTAACGGGAAGTGTACAT 89400
QY 758 tactcgcggttgatccacttttgatgcttttggatgcaattgcatgattccatcac 817
Db 89401 TACAAGTGGCTTGAAATCAACCTTTGATGCTTGGACTGCGAGTGCATGAGTTCACAC 89460
QY 818 agaagaaacaa 829
Db 89461 AGAAGGTGACAA 89472

RESULT 9
166494/c
LOCUS
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source location/Qualifiers
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN
Query Match 4.8%; Score 75.8; DB 10; Length 7218;
Best Local Similarity 5.8%; Pred. No. 3.2e-06;
Matches 23; Conservative 237; Mismatches 149; Indels 0; Gaps 0;
QY 999 gaaaatagtcagagattacatttggctactataagggcagaaatgcatgggat 1058
Db 1467 GCAAGTAGTTAAGAGATAGAGAAATTTGTCACRRRRRRRRRRRRRRRRRR 1408
QY 1059 ggatagtggttcttactttacacaaagtgagtttgcctgaaagcattatcacg 1118
Db 1407 RRR 1348
QY 1119 gaggttcaaacgcagctcaaaagtggcggtgatttcaacacattcataaacagac 1178
Db 1347 RRR 1288
QY 1179 aatcacatggtgcccgaacctccctctgtgagaggttgagaaagtgaaagaga 1238
Db 1287 RRR 1228
QY 1239 gaagagcagcatcaaaagagagttgaggagatagagaagagtagagagatgagagat 1298
Db 1227 RRR 1168
QY 1299 aaagaagtcacatttcagtaaacattgttgaggttttaaaagagctcaacagagatga 1358
Db 1167 RRR 1108
QY 1359 gagaactttaaagagagctgagcaaaagagaagaaatgaggtgtttgagat 1407
Db 1107 RRR 1059
RESULT 10
AL158158/c
LOCUS
DEFINITION Human DNA sequence from clone RP11-427L11 on chromosome 9q31.2-32, complete sequence.
ACCESSION AL158158
VERSION AL158158.14 GI:13559997
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 194835)
TITLE Direct Submission
JOURNAL Submitted (05-APR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT
On Apr 6, 2001 this sequence version replaced gi:12733508.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-427L11 is from the library RPI1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6

This sequence is the entire insert of clone RP11-427L11.

FEATURES

Source

Location/Qualifiers
 1..194835
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /map="q11.2-32"
 /clone="RP11-427L11"
 /clone_lib="RPI1-11.2"

repeat_region
 1..224
 /note="L1P repeat: matches 3..226 of consensus"
 repeat_region
 326..615
 /note="AluSg repeat: matches 1..299 of consensus"
 repeat_region
 627..772
 /note="AluSg/x repeat: matches 158..301 of consensus"
 repeat_region
 929..1168
 /note="L1M4 repeat: matches 5149..5403 of consensus"
 repeat_region
 1195..1519
 /note="AluSx repeat: matches 1..311 of consensus"
 repeat_region
 1698..1726
 /note="MER5A repeat: matches 16..44 of consensus"
 repeat_region
 1728..1868
 /note="AluSg/x repeat: matches 150..302 of consensus"
 repeat_region
 1887..1932
 /note="MER5A repeat: matches 61..106 of consensus"
 repeat_region
 1966..2242
 /note="L1ME3A repeat: matches 5614..5894 of consensus"
 misc_feature
 complement(1992..2500)
 /note="match: GSS: Em:AQ172930"
 repeat_region
 2308..2619
 /note="AluSp repeat: matches 1..313 of consensus"
 repeat_region
 2707..2822
 /note="L1ME3A repeat: matches 6017..6137 of consensus"
 repeat_region
 3336..3403
 /note="34 copies 2 mer cc 66% conserved"
 repeat_region
 3420..3725
 /note="AluSx repeat: matches 1..300 of consensus"
 repeat_region
 4319..4467
 /note="MER5B repeat: matches 4..175 of consensus"
 repeat_region
 4884..5309
 /note="L2 repeat: matches 166..616 of consensus"
 repeat_region
 5596..5870
 /note="L1MA9 repeat: matches 6027..6305 of consensus"
 repeat_region
 5871..5938
 /note="AluJ/FLAM repeat: matches 1..68 of consensus"
 repeat_region
 5967..6648
 /note="L2 repeat: matches 1284..2090 of consensus"
 repeat_region
 6649..6948
 /note="AluSg repeat: matches 1..300 of consensus"
 repeat_region
 6949..7003
 /note="L2 repeat: matches 2090..2136 of consensus"
 repeat_region
 7072..7482
 /note="L2 repeat: matches 2295..2695 of consensus"
 repeat_region
 7588..7789

misc_feature
 complement(8285..8469)
 /note="match: STS: Em:G05863"
 misc_feature
 complement(8329..8495)
 /note="match: GSS: Em:B65750"
 misc_feature
 complement(8329..8496)
 /note="match: GSS: Em:AQ227987"
 misc_feature
 complement(8339..8502)
 /note="match: GSS: Em:AQ110861"
 misc_feature
 8354..8529
 /note="match: STS: Em:G54088"
 join(8370..8532,9091..9156,15133..15192,15671..15768)
 /note="match: STS: Em:G29920"
 8591..8662
 /note="36 copies 2 mer ta 81% conserved"
 repeat_region
 9732..9893
 /note="AluJ repeat: matches 130..287 of consensus"
 repeat_region
 complement(9737..9894)
 /note="match: GSS: Em:AQ281101"
 repeat_region
 complement(9886..10119)
 /note="match: GSS: Em:B53125"
 repeat_region
 complement(9897..10091)
 /note="match: STS: Em:HSA108WE1"
 repeat_region
 complement(9907..10123)
 /note="match: GSS: Em:B17453"
 repeat_region
 9932..9967
 /note="18 copies 2 mer ta 100% conserved"
 repeat_region
 9973..10105
 /note="AluJ repeat: matches 1..133 of consensus"
 repeat_region
 10133..10392
 /note="AluSx repeat: matches 38..299 of consensus"
 repeat_region
 10183..10319
 /note="match: GSS: Em:AQ782105"
 repeat_region
 10404..10492
 /note="MER5B repeat: matches 7..97 of consensus"
 repeat_region
 10992..11078
 /note="match: GSS: Em:AQ544271"
 repeat_region
 11354..11653
 /note="AluSx repeat: matches 1..300 of consensus"
 repeat_region
 11742..12929
 /note="Tigger3b repeat: matches 2..1231 of consensus"
 repeat_region
 12987..13146
 /note="MIR repeat: matches 89..248 of consensus"
 repeat_region
 13153..13468
 /note="AluSx repeat: matches 1..312 of consensus"
 repeat_region
 13678..13848
 /note="L2 repeat: matches 2537..2703 of consensus"
 repeat_region
 14394..14609
 /note="MIR repeat: matches 12..257 of consensus"
 repeat_region
 15816..15902
 /note="match: GSS: Em:B46502"
 repeat_region
 15900..15947
 /note="24 copies 2 mer aa 79% conserved"
 repeat_region
 16257..16558
 /note="AluSg repeat: matches 1..302 of consensus"
 repeat_region
 16956..17220
 /note="AluSg repeat: matches 38..304 of consensus"
 repeat_region
 17415..17488
 /note="37 copies 2 mer aa 71% conserved"
 repeat_region
 18552..18845
 /note="AluSg repeat: matches 1..293 of consensus"
 repeat_region
 19629..19689
 /note="L2 repeat: matches 2685..2743 of consensus"
 repeat_region
 19779..21330
 /note="Cpg island"
 repeat_region
 evidence-not_experimental
 21481..21679
 /note="MIR repeat: matches 48..262 of consensus"
 repeat_region
 22591..22818
 /note="MIR repeat: matches 23..258 of consensus"
 repeat_region
 complement(23067..23458)
 /note="match: GSS: Em:B43385"
 repeat_region
 complement(23074..23458)

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 45698: contig of 45698 bp in length
 * 45699: gap of unknown length
 * 45799: contig of 26112 bp in length
 * 71910: contig of unknown length
 * 72010: gap of unknown length
 * 72011: gap of 22924 bp in length
 * 94934: contig of 22924 bp in length
 * 94935: gap of unknown length
 * 113935: contig of 18900 bp in length
 * 114034: gap of unknown length
 * 126397: contig of 12363 bp in length
 * 126498: gap of unknown length
 * 137932: contig of 11435 bp in length
 * 138032: gap of unknown length
 * 149539: contig of 11507 bp in length
 * 149639: gap of unknown length
 * 161975: contig of 12336 bp in length
 * 162075: gap of unknown length
 * 161976: contig of 9627 bp in length
 * 171702: contig of unknown length
 * 171802: gap of unknown length
 * 171803: contig of 7872 bp in length
 * 179675: gap of unknown length
 * 179775: contig of 6064 bp in length
 * 185938: gap of unknown length
 * 185939: contig of 4942 bp in length
 * 190881: gap of unknown length
 * 190981: contig of 2417 bp in length
 * 193397: gap of unknown length
 * 193398: contig of 1118 bp in length.
 * 193498: Location/Qualifiers

FEATURES

source

1..194615
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-438D8"

BASE COUNT 57084 a 39768 c 39505 g 56942 t 1316 others
 ORIGIN

Query Match 3.4%; Score 54; DB 61; Length 194615;
 Best Local Similarity 52.7%; Pred. No. 0.18;
 Matches 117; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
 QY 1217 ggagaagaagtggagaaggaaggaaggaagcgcataaaagaagtggagatagaaga 1276
 Db 85526 GAAGAAGAAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGA 85585
 QY 1277 agaagttagaaggtgagagaagaagtcaccttattcagtaaaactgtttgaagttt 1336
 Db 85586 AGAAG 85645
 QY 1337 taagaagctccacgagatgaagagaactctttaagagagctgagcaagaagaatgga 1396
 Db 85646 AGAAG 85705
 QY 1397 tgttttgatgactaaatgaagcaactgaggttagaaaa 1438
 Db 85706 GGAGAGGAGGAG 85747

RESULT 13

AC016152 178189 bp DNA HTG 17-APR-2001
 LOCUS
 DEFINITION Homo sapiens chromosome 12 clone RP11-1006M13, WORKING DRAFT

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS

SEQUENCE, 6 unordered pieces.

AC016152 GI:13654320
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 178189)

Musny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
 Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
 Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Loulseghe,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
 Ogih,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
 Sisson,I., Sodergren,E., Sonaite,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
 and Gibbs,R.

Direct Submission

Unpublished
 2 (bases 1 to 178189)

Worley,K.C.

Direct Submission

Submitted (23-NOV-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On Apr 17, 2001 this sequence version replaced gi:13162462.

COMMENT

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HMRV
 Center clone name: RP11-1006M13
 ----- Summary Statistics
 Sequencing vector: Plasmid: M77789
 Sequencing vector: M13: L08821
 Chemistry: Dye-primer Bodypy: 23% of reads
 Chemistry: Dye-terminator Big Dye: 77% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 174861 bases at least Q40
 Consensus quality: 176412 bases at least Q30

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2001, 02:55:37 ; Search time 313.48 Seconds
(without alignments)
3182.773 Million cell updates/sec

Title: US-09-075-375A-3

Perfect score: 1589

Sequence: 1 tatttttcagtggttcagt.....tcttttcagaagcaaaaaa 1589

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601:*

- 1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1587.4	99.9	1589	18	AAT66242 Tobacco violaxanthin
2	647	40.7	1981	18	AAT66241 Romaine lettuce vi
3	622.6	39.2	1555	18	AAT66243 Arabidopsis violax
C 4	125	7.9	936	22	AAF58252 Oligonucleotide D1
C 5	125	7.9	936	22	AAF58254 Oligonucleotide D1
C 6	125	7.9	936	22	AAF58257 Oligonucleotide D1
C 7	125	7.9	936	22	AAF58259 Oligonucleotide D2
C 8	125	7.9	936	22	AAF58262 Oligonucleotide D2
C 9	125	7.9	938	22	AAF58255 Oligonucleotide D1
10	121.2	7.6	936	22	AAF58252 Oligonucleotide D1
11	121.2	7.6	936	22	AAF58254 Oligonucleotide D1

12	121.2	7.6	936	22	AAF58257 Oligonucleotide D1
13	121.2	7.6	936	22	AAF58259 Oligonucleotide D2
14	121.2	7.6	936	22	AAF58262 Oligonucleotide D2
15	121.2	7.6	938	22	AAF58255 Oligonucleotide D1
16	52.4	3.3	1686	16	AAQ87587 DNA encoding Leuco
17	52	3.3	4000	18	AAQ87587 Mannose-1-phosphat
C 18	51.2	3.2	244	22	AAF58238 Oligonucleotide D1
19	50.6	3.2	3211	18	AAT89346 Human p160 cDNA 16
20	50.6	3.2	3901	18	AAT89345 Human p160 cDNA 16
21	49	3.1	244	22	AAF58238 Oligonucleotide D1
22	48.6	3.1	4677	21	AAAT70259 Plasmodium falcipa
23	48.6	3.1	5361	18	AAT78868 P. falciparum live
24	48.6	3.1	6152	18	AAT78867 P. falciparum live
25	47.4	3.0	3579	21	AAAT00099 Plasmodium falcipa
26	46.6	2.9	11011	21	AAC68252 B. burgdorferi tox
27	46.6	2.9	910715	20	AAZ20248 Borrelia burgdorfe
28	45	2.8	3399	17	AAT05868 Chicken leucocytos
29	43.8	2.8	4590	7	AA60472 Sequence encoding
C 30	43.6	2.7	8952	19	AAV62130 HSV-2 strain SB5 C
31	43	2.7	979	20	AAV61816 B. burgdorferi ant
32	43	2.7	1083	20	AAV61815 B. burgdorferi ant
33	43	2.7	2487	20	AAZ20298 Borrelia burgdorfe
34	42.8	2.7	49999	20	AAZ23891 Murine LOBO genom
35	42.8	2.7	49999	20	AAZ23896 Murine LOBO homolo
C 36	42.4	2.7	1236	21	AAZ02163 Human colon cancer
37	42.2	2.7	1380	21	AAZ34650 Human growth facto
C 38	42	2.6	6789	22	AAV63436 Murine CD39-L4 gen
39	42	2.6	29392	19	AAV15422 Mouse poly Ig rece
C 40	41.6	2.6	3250	21	AAAG4292 Nucleotide sequenc
C 41	41.6	2.6	3795	20	AAZ25348 Mouse DRM nucleoti
42	41.2	2.6	5940	21	AAAT0105 Plasmodium falcipa
43	40.8	2.6	5394	16	AAT00872 Murine mc26 struct
C 44	40.8	2.6	5394	21	AAZ86916 Mouse mc26 gene.
C 45	40.2	2.5	2010	10	AAAN91379 Intron 1 from huma

ALIGNMENTS

RESULT 1

AAT66242

ID AAT66242 standard; cdna; 1589 BP.

XX AC AAT66242;

XX DT 28-JUL-1997 (first entry)

XX DE Tobacco violaxanthin de-epoxidase cdna.

XX KW violaxanthin de-epoxidase; VDE; light; photosensitivity; photoprotection; transgenic plant; zeaxanthin; antheraxanthin;

XX KW xanthophyll; lettuce; ss.

XX OS Nicotiana tabacum L. cv. xanthi.

XX FH Key Location/Qualifiers

XX CDS 42...1478

XX FT /*tag= a

XX FT transit_peptide 42...443

XX FT /*tag= b

XX FT mat_peptide 444...1475

XX FT /*tag= c

XX FT WO9717447-A2.

XX PN 15-MAY-1997.

XX PD 07-NOV-1996; 96WO-US18291.

XX PF 06-AUG-1996; 96US-0023502.

XX PR 07-NOV-1995; 95US-0006315.

XX XX (CALJ) CALGENE INC.

XX Bugos RC, Rockholm DC, Yamamoto HY;
 XX WPI; 1997-281036/25.
 DR P-PSDB; AAW09875.
 XX DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light
 XX Disclosure; Fig 2; 41pp; English.
 XX A cDNA clone (AAW66242) codes for tobacco violaxanthin de-epoxidase
 CC (VDE) (AAW09875), an enzyme that catalyses the de-epoxidation of
 CC violaxanthin to zeaxanthin and antheraxanthin. VDE nucleic acids
 CC (see also AAT66241, AAT66243), in sense or antisense orientation, can
 CC be used in genetic constructs, *prel.* also contg. a plastid
 CC translocation sequence, to modify VDE levels in plants. Increased
 CC levels result in the plant being tolerant of increased light and
 CC therefore more productive and/or more resistant to disease.
 CC Underexpression of VDE increases photosynthetic efficiency under
 CC low light. The photosensitivity of a range of crops, trees and
 CC ornamentals can be modified.
 XX SQ Sequence 1589 BP; 500 A; 261 C; 367 G; 461 T; 0 other;

Query Match 99.9%; Score 1587.4; DB 18; Length 1589;
 Best Local Similarity 99.9%; Pred. NO. 0;
 Matches 1588; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tattttcatgagtttcagttggtgtaatacagtttgaagaatggtcttgccttcatt 60
 DB 1 tattttcatgagtttcagttggtgtaatacagtttgaagaatggtcttgccttcatt 60

QY 61 caaatcttctggccaccatgaaccatcaaatattattattgttgggtcaaaagtctccggtc 120
 DB 61 caaatcttctggccaccatgaaccatcaaatattattattgttgggtcaaaagtctccggtc 120

QY 121 ataaaggtttagctgggttgggaagattactttgttagtagtagctgtagcaaaaattt 180
 DB 121 ataaaggtttagctgggttgggaagattactttgttagtagtagctgtagcaaaaattt 180

QY 181 gtccagcagcaggtacactagatactttcgaataatctcctagaatatgctgtggttgg 240
 DB 181 gtccagcagcaggtacactagatactttcgaataatctcctagaatatgctgtggttgg 240

QY 241 attcaagaggtctgcaactattctcacggaacacacaaatctctcccgacacatagca 300
 DB 241 attcaagaggtctgcaactattctcacggaacacacaaatctctcccgacacatagca 300

QY 301 ttaaccagaatgtacctaagggaatttcaggatgcgaatttccaaaagatgagctttga 360
 DB 301 ttaaccagaatgtacctaagggaatttcaggatgcgaatttccaaaagatgagctttga 360

QY 361 tggtttggagaataatgggccaatttgcgaacacagcaaatgttagctatattcatttgc 420
 DB 361 tggtttggagaataatgggccaatttgcgaacacagcaaatgttagctatattcatttgc 420

QY 421 cagttgcttcaaaagctgtagcgggttgatgctctcaagacttgccttactgaaag 480
 DB 421 cagttgcttcaaaagctgtagcgggttgatgctctcaagacttgccttactgaaag 480

QY 481 agtgcaggttagagcttggaagtgcatttcgaacctgcattgcagcctaattgttgcct 540
 DB 481 agtgcaggttagagcttggaagtgcatttcgaacctgcattgcagcctaattgttgcct 540

QY 541 gtctccagacttgcacaatagactgcagcaacggaatgtcagataaaatgtgtgatt 600
 DB 541 gtctccagacttgcacaatagactgcagcaacggaatgtcagataaaatgtgtgatt 600

QY 601 tggttgaaaacagctgctgtagcaggttcaatgagtggtgcagctctcccgaaagaaatgtg 660
 DB 601 tggttgaaaacagctgctgtagcaggttcaatgagtggtgcagctctcccgaaagaaatgtg 660

RESULT 2

AAT66241

ID AAT66241 standard; cDNA; 1981 BP.

XX AAT66241;

DT 28-JUL-1997 (first entry)

QY 661 taccctgtaaatctgattgttggaacttctctgtacccagtgatccagtggtcttgcaga 720
 DB 661 taccctgtaaatctgattgttggaacttctctgtacccagtgatccagtggtcttgcaga 720

QY 721 agtttgacatgaagaatttttagcgggaaatggttcattactcoggttgaatcccaatt 780
 DB 721 agtttgacatgaagaatttttagcgggaaatggttcattactcoggttgaatcccaatt 780

QY 781 ttgatgtttttgattgccaatgtgcattgtccatacagaagaacaaacaaacttggggga 840
 DB 781 ttgatgtttttgattgccaatgtgcattgtccatacagaagaacaaacaaacttggggga 840

QY 841 atttatcttgagaaatcgtacacactgattgagagattttttactcgtacgcggtgcaga 900
 DB 841 atttatcttgagaaatcgtacacactgattgagagattttttactcgtacgcggtgcaga 900

QY 901 aattcgtgcaagatcccaaatgtatccggggatctctcaaatcatgataatgagtatcttc 960
 DB 901 aattcgtgcaagatcccaaatgtatccggggatctctcaaatcatgataatgagtatcttc 960

QY 961 tctaccaagatgactggttatattttgtcatcccaagtagaaaatagtcacagagattaca 1020
 DB 961 tctaccaagatgactggttatattttgtcatcccaagtagaaaatagtcacagagattaca 1020

QY 1021 tatttgttactataaagggcagaaatgatgcattgggtggtgatatgttgggttctgtacttt 1080
 DB 1021 tatttgttactataaagggcagaaatgatgcattgggtggtgatatgttgggttctgtacttt 1080

QY 1081 acacaagaatgcagtttgcctgaaagcattatcccgaggttgcgaacccgcagctcaaa 1140
 DB 1081 acacaagaatgcagtttgcctgaaagcattatcccgaggttgcgaacccgcagctcaaa 1140

QY 1141 aagttggcgtgatttcaacacattcaaaaacagacaatacatatgtggcctgaaacctc 1200
 DB 1141 aagttggcgtgatttcaacacattcaaaaacagacaatacatatgtggcctgaaacctc 1200

QY 1201 cccttgttgagaggttggaagaagaatgggaagagagaagagacatcataaaagaag 1260
 DB 1201 cccttgttgagaggttggaagaagaatgggaagagagaagagacatcataaaagaag 1260

QY 1261 ttgaggagatagaagaagaatagagagtagagagtagagagataaagaagtcaccttaccga 1320
 DB 1261 ttgaggagatagaagaagaatagagagtagagagtagagagataaagaagtcaccttaccga 1320

QY 1321 aactgtttgaaggttttaagagctcccaacgagatgaagagaacttcttaagagagctga 1380
 DB 1321 aactgtttgaaggttttaagagctcccaacgagatgaagagaacttcttaagagagctga 1380

QY 1381 gcaagaagaagaatggatgttttggatggacttaaaatgggaagcaactgaggtgagaaaac 1440
 DB 1381 gcaagaagaagaatggatgttttggatggacttaaaatgggaagcaactgaggtgagaaaac 1440

QY 1441 tttttggcgtctttaccataagaataaagtaagtaatttttttaaaactatcaaat 1500
 DB 1441 tttttggcgtctttaccataagaataaagtaagtaatttttttaaaactatcaaat 1500

QY 1501 atatactacatgtagttagttgatttgcctggaatagattgcttatacatca 1560
 DB 1501 atatactacatgtagttagttgatttgcctggaatagattgcttatacatca 1560

QY 1561 tggattgcttcttttcagaagaacaaaaa 1589
 DB 1561 tggattgcttcttttcagaagaacaaaaa 1589

XX Romaine lettuce violaxanthin de-epoxidase cDNA.
 DE
 XX
 KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 KW xanthophyll; lettuce; ss.
 XX
 OS Lactuca sativa L. cv. romaine.
 XX
 FH Key Location/Qualifiers
 FT misc_difference 26..29
 FT /tag= a
 FT /note= "bases 26-29 are illegible in Fig 1"
 FT misc_difference 66..72
 FT /tag= b
 FT /note= "bases 66-72 are illegible in Fig 1"
 FT misc_difference 105..110
 FT /tag= c
 FT /note= "bases 105-110 are illegible in Fig 1"
 FT misc_difference 147..149
 FT /tag= d
 FT /note= "bases 147-149 are illegible in Fig 1"
 FT misc_difference 186..189
 FT /tag= e
 FT /note= "bases 186-189 are illegible in Fig 1"
 FT misc_difference 226..227
 FT /tag= f
 FT /note= "bases 226-227 are illegible in Fig 1"
 FT CDS
 FT 235..1656
 FT /tag= g
 FT transit_peptide 235..609
 FT /tag= h
 FT mat_peptide 610..1653
 FT /tag= i
 XX
 XX WO9717447-A2.
 PN
 XX
 PD 15-MAY-1997. 96WO-US18291.
 PF 07-NOV-1996; 96WO-US18291.
 XX
 PR 06-AUG-1996; 96US-0023502.
 PR 07-NOV-1995; 95US-0006315.
 XX
 PA (CALJ) CALGENE INC.
 XX
 PI Bugos RC, Rockholm DC, Yamamoto HV;
 XX
 DR WPI; 1997-281036/25.
 XX P-PSDB; AAW09874.
 XX
 PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light
 XX
 PS Claim 3; Fig 1; 4lpp; English.
 XX
 CC A cDNA clone (AAT66241) codes for the 55 kDa violaxanthin de-epoxidase
 CC (VDE) (AAW09874) of romaine lettuce. VDE was purified from romaine
 CC lettuce chloroplasts and 2 tryptic peptides were used to develop
 CC primers (see also AAT66244-45), which amplified a partial VDE
 CC sequence. The amplified sequence was then used to screen a lettuce
 CC cDNA library, and the 1981 bp DNA sequence was identified. VDE
 CC nucleic acids (see also AAT66242-43), in sense or antisense
 CC orientation, can be used in genetic constructs, pref. also contg. a
 CC plastid translocation sequence, to modify VDE levels in plants.
 CC Increased levels result in the plant being tolerant of increased
 CC light and therefore more productive and/or more resistant to
 CC disease. Underexpression of VDE increases photosynthetic
 CC efficiency under low light. The photosensitivity of a range of
 CC crops, trees and ornamentals can be modified.
 XX
 SQ Sequence 1981 BP; 608 A; 337 C; 433 G; 577 T; 26 other;

Query Match 40.7%; Score 647; DB 18; Length 1981;
 Best Local Similarity 71.9%; Pred. No. 3.9e-156;
 Matches 864; Conservative 0; Mismatches 325; Indels 12; Gaps 1;
 QY 317 taagggaattcaggatgcacaaattccaaaagatgtagctttgatggtttgggagaatg 376
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 483 tgaggaaatacaaaatttgatctcaaaaggggcatgacttgattcttgaagaacatg 542
 QY 377 gggccaaattgccaacacagcaattgtactatattcattttgtcagttgcttcaaaagc 436
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 543 gagacaattcatacaatttgctatgtttgttcacattttgtattgttcttagagt 602
 QY 437 tgatgcggttgatctcgaagacttgactgttactgaagaagatgcaggttagagt 496
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 603 tgatgcggttgatctcgaagacttgactgttactgaagaagatgcaggttagagt 662
 QY 497 tgcgaagtgcatttcgaacctgcatgtgcagctaatgttgcctgtctccagacttgcaa 556
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 663 tgcgaagtgcatttcgaacctgcatgtgcagctaatgttgcctgtctccagacttgcaa 722
 QY 557 caatagacctgcgaacacggaattgtcagataaaattgtgtgtttgttgaacacagtgt 616
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 723 caatgcctgcagagaccgaattgcagataaaattgtgtgtttgttgaacacagtgt 782
 QY 617 cgtagacagattcaatgagtgtcagctcccgaaagaaatgtgtacotcgttaactcga 676
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 783 ggtggaccattcaacgagtgtgcggttcccgaaagaaatgtgtgcccggaaatcgga 842
 QY 677 tgtgtgactttcctgtacactgccaggtgtcttctccagaatttgcacatgaaga 736
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 843 tgtgtgtgaattcccggtttccggatgtgaatgcagtggttcaaaattttaacatgaaga 902
 QY 737 ttttagcgggaattgttctactcgcggtttggaatcccaacttttgatgttttgattg 796
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 903 ctttagtgggaattgttatatacaaatgttttaaatcccaacttttgatgttcattgattg 962
 QY 797 ccaattgcatgattccatcacagaaagaaacaaactttgggaaatttatcttggagaat 856
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 963 tcaactcatgattctcattgataaactgttgggaactttggaacttaactgacatc 1022
 QY 857 acgtacacctgatggaggatttttttactcgcagcgggtgcaaaaattcgtgcaagatcc 916
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1023 aaaaactttgagtggttttttactcgcagcgtgtgtgcaaacatttgttcaagatcc 1082
 QY 917 aaagtatccgggatactctacaatcatgataatgagtattcttctcacaagatgactg 976
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1083 agatcttctcgtggagcatttataatcatgacaaatgagtgttcttcaactcaagatgactg 1142
 QY 977 gtatatatttgcataccaaagttagaaaatagtcagagagattacatatatttgtactataa 1036
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1143 gtacattattcttcccaatcgaaaacacccgcagattcacacattcgttactacacg 1202
 QY 1037 gggcagaattgatggatggatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1096
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1203 aggtcgaaacgacgctggatggatgacggtgggttcggtgatttaccacccgaagtcgcac 1262
 QY 1097 ttgtcctgaagcattatataccggaggttgcaaccgcagctcaaaaattgtggcggtgattt 1156
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1263 actcccggaatgcacatcccaaacatcaaaaagcagcacaatccggtgggtcgagactt 1322
 QY 1157 caacacattcataaaaaacagacaatacatgtgtgcctgaacctccccttgtttgagaggtt 1216
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1323 taacaatttctaaacacccgacataatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1382
 QY 1217 ggagaagaagtggagaagggaggaagggacgacatcataaaagaagtgtgagagatagaaga 1276
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1383 tgagaacacagcgaaggggagaggt 1442
 QY 1277 agaagta-----gagaaggtgagagataaaagtaagtcaccttatttcagtaaac 1324
 Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1443 agaggttgaaaaagaggttgagaaggttagagatactagatgactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1502

QY 1325 gttgaaggttttaagagagctccaacgagatgaagagaaacttcttaagagagctgagcaa 1384
 Db 1503 gcttgaaggttttaagagagctccaacgagatgaagagaaacttcttaagagagctgagtaa 1562
 QY 1385 agaagaaatgatttttgatgagcttaaaatggaacactgaggtagaaaaactttt 1444
 Db 1563 agaagaaatgatttttgatgagcttaaaatggaacactgaggtagaaaaactttt 1622
 QY 1445 tgggcgtgctttaccataaggaataaggttaagtatttttaaaactatacaatata 1504
 Db 1623 tgggcgtgctttaccataaggaataaggttaagtatttttaaaactatacaatata 1682
 QY 1505 a 1505
 Db 1683 a 1683

RESULT 3

AA166243 ID AAT66243 standard; cDNA; 1555 BP.

AC AAT66243;

XX 28-JUL-1997 (first entry)

XX Arabidopsis violaxanthin de-epoxidase cDNA.

DE Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 KW xanthophyll; lettuce; ss.

XX Arabidopsis thaliana cv. columbia.

XX Key Location/Qualifiers
 FH 45..1433
 CDS /*tag= a

FT transit_peptide 45..383
 FT /*tag= b

FT mat_peptide 384..1430
 FT /*tag= c

XX WO9717447-A2.

XX 15-MAY-1997.

XX 07-NOV-1996; 96WO-US18291.

XX 06-AUG-1996; 96US-0023502.

XX 07-NOV-1995; 95US-0006315.

XX (CALJ) CALGENE INC.

XX Bugos RC, Rockholm DC, Yamamoto HY;

XX WPI; 1997-281036/25.

XX P-PSDB; AAW09876.

XX DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light

XX Disclosure; Fig 3; 41pp; English.

XX A cDNA clone (AAT66243) codes for Arabidopsis violaxanthin
 CC de-epoxidase (VDE) (AAW09876), an enzyme that catalyses the
 CC de-epoxidation of violaxanthin to zeaxanthin and antheraxanthin.
 CC VDE nucleic acids (see also AAT66241-42), in sense or antisense
 CC orientation, can be used in genetic constructs, pref. also contg. a
 CC plastid translocation sequence, to modify VDE levels in plants.
 CC Increased levels result in the plant being tolerant of increased
 CC light and therefore more productive and/or more resistant to
 CC disease. Underexpression of VDE increases photosynthetic
 CC efficiency under low light. The photosensitivity of a range of
 CC crops, trees and ornamentals can be modified.

XX Sequence 1555 BP; 485 A; 286 C; 375 G; 409 T; 0 other;
 SQ

Query Match 39.2%; Score 622.6; DB 18; Length 1555;
 Best Local Similarity 74.3%; Pred. No. 6.3e-150;
 Matches 819; Conservative 0; Mismatches 269; Indels 15; Gaps 2;

QY 424 ttgcttcaaaagctgagcggttgatgctctcaagacttgaacttgaactgaagaagt 483
 Db 364 ttgttccatctgcagagcagtgatgacacttaaaacttggcatgcttattgaaggat 423
 QY 484 gcaggttagagcttgcgaagtgcatttcgaacctgcattgcagctaaagtgtcctgtc 543
 Db 424 gcaggttagagcttgcgaagtgcatttcgaacctgcattgcagctaaagtgtcctgtc 483
 QY 544 tccagacttgcacaatagacctgacgaacggaatgtcagataaaatgtgtgattgt 603
 Db 484 tccagacttgcacaatagacctgacgaacggaatgtcagataaaatgtgtgattgt 543
 QY 604 ttgaaacagtgctgtagacgagttcaatgagtgagtgctcccgaaagaaatgtgtac 663
 Db 544 ttgaaacagtgctgtagacgagttcaatgagtgagtgctcccgaaagaaatgtgtac 603
 QY 664 ctgctaaatctgagtggtgagcttctctgactgacccagtgcttctgtccgaagt 723
 Db 604 ctgctaaatctgagtggtgagcttctctgactgacccagtgcttctgtccgaagt 663
 QY 724 ttgcatgaaagatttagcgggaatggttcattactcgggttgaatcccaactttg 783
 Db 664 ttgcatgaaagatttagcgggaatggttcattactcgggttgaatcccaactttg 723
 QY 784 atgcttttgattgccaattgcatgagttccatatacagaa---gaaacaaactgtgtgga 840
 Db 724 atgcttttgattgccaattgcatgagttccatatacagaa---gaaacaaactgtgtgga 783
 QY 841 atttacttggagaatcgtacacactgagtgaggtattttactcgcagcggtgcaaa 900
 Db 784 atttacttggagaatcgtacacactgagtgaggtattttactcgcagcggtgcaaa 843
 QY 901 aattctgcaagatccaaagtatccggggaactctcaatcatgataatgagtattctc 960
 Db 841 aattctgcaagatccaaagtatccggggaactctcaatcatgataatgagtattctc 903
 QY 961 tctaccaagatgactggtatattttgtccatccaaagttagaaaatagtcacagagattaca 1020
 Db 901 tctaccaagatgactggtatattttgtccatccaaagttagaaaatagtcacagagattaca 963
 QY 1021 tatttgtactataaaggcggaagaatgcatggtggatggatgtgtgtctgtacttt 1080
 Db 961 tatttgtactataaaggcggaagaatgcatggtggatggatgtgtgtctgtacttt 1023
 QY 1081 acacaagaagtgcagttttgctgaaagcattataccggaggttgcacacccgagactcaaa 1140
 Db 1021 acacaagaagtgcagttttgctgaaagcattataccggaggttgcacacccgagactcaaa 1083
 QY 1141 aagttggcggtgatttcaaacattataaaacagacaatacatatgtggccctgaacctc 1200
 Db 1081 aagttggcggtgatttcaaacattataaaacagacaatacatatgtggccctgaacctc 1143
 QY 1201 ccttctgtgagaggttggagagaagaagtggagaagagagagacgactcataaaagaag 1260
 Db 1141 ccttctgtgagaggttggagagaagaagtggagaagagagagacgactcataaaagaag 1203
 QY 1261 ttggagagatagaagaagaatgagagaaggtgagagataaagaag-----tca 1308
 Db 1201 ttggagagatagaagaagaatgagagaaggtgagagataaagaag-----tca 1263
 QY 1309 ccttattcaglaaactgttgaaggttttaagagcttccacagcagatgaagagacttct 1368
 Db 1261 ccttattcaglaaactgttgaaggttttaagagcttccacagcagatgaagagacttct 1323
 QY 1369 taagagagctgagcaagaagaagaatgagtgtttggtgagacttaaaatggaagcaactg 1428


```

Db 1324 tgagagagtttaagtaagaagagatggagttttttgatacagatcaaaatggaagcaagtg 1383
QY 1429 aggtagaacaaactttttggcggtgtcttaccacaataaggaaattaaagttaatttttaa 1488
Db 1384 aggttgaacaaattgtttgggaagctttgccaatcggaaggtcaggtagaacaagaac 1443
QY 1489 aactatcaacataatactacat 1511
Db 1444 caccattgtgtacaactatat 1466

RESULT 4
ID AAF58252/c
XX AAF58252 standard; DNA; 936 BP.
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
SQ WPI; 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in
hybridization assays, e.g. for genotyping, allowing repeat analyses on
a single surface

Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 7.9%; Score 125; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 1.5e-22;
Matches 5; Conservative 489; Mismatches 289; Indels 0; Gaps 0;

QY 799 aattgcatgagttcattacagaagaacaaacttgtgggaatttattctctggagaataac 858
Db 785 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 726
QY 859 gtacacctgatggaggattttttactcgcagcggtgcacaaatctgtgcaagatccaa 918
Db 725 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 666
QY 919 agtatccggggatactctacaatcatgataatgagtatctctctaccaagatgactggt 978
Db 665 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 606

```

```

QY 979 atattttgcatccaaatagataaaatagtcacagaggattacatatattgtgtactataagg 1038
Db 605 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 546
QY 1039 gcagaaatgatcgatgggatgatatgtgtctgtactttcacacaagaagtgcagttt 1098
Db 545 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 486
QY 1099 tgcctgaaagcattataccggagttgcaaacgcgcagctcacaagaagtgtggcggtatttca 1158
Db 485 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 426
QY 1159 acacattcataaaacagacaatacatatgtggccctggaacctccctctgtgtgaggggttg 1218
Db 425 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 366
QY 1219 agaagaagtggaaaggaaggaagggacgacatcataaaagaagtttgaggagatagaagaag 1278
Db 365 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 306
QY 1279 aagtagaagagtgagagataaagaagtcaccttattcagtaaaactgtttgaaggtttta 1338
Db 305 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 246
QY 1339 aagagctccaaacgagatgaagaacacttcttaagagagctgagcaagaagaatgatg 1398
Db 245 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 186
QY 1399 ttttgatggacttaaaatggaagcaactgaggtagaaaactttttggcggtgctttac 1458
Db 185 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 126
QY 1459 caataaggaataagtgtaagtatttttaaaactatcaacatatatactacatgtatagt 1518
Db 125 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 66
QY 1519 tgtattgatctctttgcttgatagattgtcttatacatcatgtattgtctcttttca 1578
Db 65 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 6
QY 1579 gaa 1581
Db 5 WWW 3

RESULT 5
ID AAF58254/c
XX AAF58254 standard; DNA; 936 BP.
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;

```


Db 366 www.....wgcctaw..... 425
 QY 1234 aaggagaagcagatcataaagaagttgagagatagaagaagtagaaggtga 1293
 Db 426 www..... 485
 QY 1294 gagataaagaagtcaccttattcagtaaaactgtttgaaggttttaaaagagctcccaacgag 1353
 Db 486 www..... 545
 QY 1354 atgaagagaactcttaagaagcctgagcaagaagaatgatgttttgatgagactta 1413
 Db 546 www..... 605
 QY 1414 aaatgaagaactgaggtagaaaacttttggcgctgttaccaataaaggaataa 1473
 Db 606 www..... 665
 QY 1474 ggtaagtatttttaaaactatcaacatatatactacatgtatgtttgtattgtcttt 1533
 Db 666 www..... 725
 QY 1534 tgctggaatagattgctttacatcatgtattgtcttttttcagaagaacaaaaa 1589
 Db 726 www..... 781

RESULT 11

AAF58254

ID AAF58254 standard; DNA; 936 BP.

XX AC

XX AAF58254;

XX 24-APR-2001 (first entry)

XX DE Oligonucleotide D1875.

XX XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX OS

XX Synthetic.

XX PN WO200107665-A2.

XX PD

XX 01-FEB-2001.

XX PF 26-JUL-2000; 2000WO-US20476.

XX PR 26-JUL-1999; 99US-0145695.

XX PR 17-MAR-2000; 2000US-0190259.

XX XX (CLIN-) CLINICAL MICRO SENSORS INC.

XX PA Umek RM;

XX PI WPI; 2001-159728/16.

XX DR

XX XX Nucleic acids containing electron-transfer group, useful as labels in

XX PT hybridization assays, e.g. for genotyping, allowing repeat analyses on

XX PT a single surface

XX XX

XX PS Example 6; Page 127; 159pp; English.

XX CC The present invention relates to a composition comprising two nucleic

XX CC acids each containing an electron-transfer group (ETM) having

XX CC different redox potentials. The invention is used for electronic

XX CC detection of nucleic acids, especially of substitutions (mismatches)

XX CC and single-nucleotide polymorphisms, e.g. for genotyping,

XX CC monitoring gene expression.

XX XX

XX OS Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

XX XX

XX PN WO200107665-A2.

Query Match 7.6%; Score 121.2; DB 22; Length 936;
 Best Local Similarity 0.4%; Pred. No. 1.4e-21;
 Matches 3; Conservative 485; Mismatches 288; Indels 0; Gaps 0;

QY 814 atacagagaagaaacaaactgtgggaatttatcttgagaaatactacacctgatggag 873
 Db 6 www..... 65
 QY 874 gatttttactcgatcagcggtgcacaaatctgtcgaagatcccaagatcccggggagac 933
 Db 66 www..... 125
 QY 934 tctacaatcatgataatgagtatctctctacaaagatgactgttatatttgcaccca 993
 Db 126 www..... 185
 QY 994 aagttagaaaatagtcacagaggattacatatatttggctactataagggcagaaatgatcat 1053
 Db 186 www..... 245
 QY 1054 gggatgatatgtgttctgtactttacacaaagtcagtttttcctgaaagcatta 1113
 Db 246 www..... 305
 QY 1114 taccggagttgcacacccgacgctcaaaaagttggcgctgatttcaacacattcataaaaa 1173
 Db 306 www..... 365
 QY 1174 cagacaatcatgtggccctgacacctcccttgttgagaggttgagaagaagtggaag 1233
 Db 366 www..... 425
 QY 1234 aaggagaagcagatcataaagaagttgagggagatagaagaagaagtagagaagtgga 1293
 Db 426 www..... 485
 QY 1294 gagataaagaagtcaccttattcagtaaaactgtttgaggttttaaaagagctccacag 1353
 Db 486 www..... 545
 QY 1354 atgaagagaactctttaaagagcgtgagcaagaagaatggtgttttgatgagactta 1413
 Db 546 www..... 605
 QY 1414 aaatggaagcaactgaggtagaaaaacttttggcgctgtttaccataaaggaaataa 1473
 Db 606 www..... 665
 QY 1474 ggtagatatttttaaaactatcaacatatatactacatgtatagttgtattgtatttt 1533
 Db 666 www..... 725
 QY 1534 tgctggaatagattgctttacatcatgtattgtcttttttcagaagaacaaaaa 1589
 Db 726 www..... 781

RESULT 12

AAF58257

ID AAF58257 standard; DNA; 936 BP.

XX AC

XX AAF58257;

XX XX

XX DT 24-APR-2001 (first entry)

XX DE Oligonucleotide D1954.

XX XX Electron-transfer group; ETM; mismatch; genotyping;

KW gene expression; ss.

XX OS

XX Synthetic.

XX XX

XX PN WO200107665-A2.

```
XX 01-FEB-2001.
PD
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 7.6%; Score 121.2; DB 22; Length 936;
Best Local Similarity 0.4%; Pred. No. 1.4e-21;
Matches 3; Conservative 485; Mismatches 288; Indels 0; Gaps 0;

Qy 814 atacagaagaacaaactgtgggaatttattcttgagaatacgtcacctgatggag 873
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 874 gatctttactcagcagcggtgcaaaatcgtgcaagatccaaatattccggggatcac 933
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 934 tctcaatcatgataatgagtatcttctaccagaagtgactgtgtatttgcctgaaagcatta 1113
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 126 www. 125
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 185 www. 185
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1053 aagtagaaaaatagccagagattacatttctgtactataaggcagaaatgatgat 1053
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1054 gggatgatattgttctgtactttacacaagaagtcagtttgcctgaaagcatta 1113
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1114 taccggagttgcaaacccagcagctcaaaaagtggcgtgatttcaacacattcattataaaa 1173
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1174 cagacaatacatgtgcccctgaacctccctctgttgagaggttggaagaagtggaag 1233
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1234 aaggagaagagcagatcataaaagaagttgagagatagaagaagtagaagaagtgta 1293
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1294 gagataagaagtcacctatttcagtaaaactgtttgaggttttaagagctccaacag 1353
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1354 atgaagaagaacttcttaagagagctgagcaagaagaagaaatgagtttttgatggaactta 1413
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 546 www. 605
```

```
Qy 1414 aatggaagcaactgaggtagaaaaactttttggcggtgtctttaccataaagaaaattaa 1473
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1474 ggtaagtatttttaaaactatcaacatacatatcatgtatgtattgtattgtattcttt 1533
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1534 tgcctggaatagattgtcttatacatcatgtattgtcttttttcagaagaacaaaaa 1589
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 726 www. 781
```

RESULT 13

AAF58259
ID AAF58259 standard; DNA; 936 BP.

XX AC AAF58259;

XX DT 24-APR-2001 (first entry)

XX DE Oligonucleotide D2004.

XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.

XX OS Synthetic.

XX PN WO200107665-A2.

XX PD 01-FEB-2001.

XX PF 26-JUL-2000; 2000WO-US20476.

XX PR 26-JUL-1999; 99US-0145695.

XX PR 17-MAR-2000; 2000US-0190259.

XX PA (CLIN-) CLINICAL MICRO SENSORS INC.

XX PI Umek RM;

XX DR WPI; 2001-159728/16.

XX PT Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface

XX PS Example 6; Page 128; 159pp; English.

XX CC The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.

XX SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 7.6%; Score 121.2; DB 22; Length 936;
Best Local Similarity 0.4%; Pred. No. 1.4e-21;

Matches 3; Conservative 485; Mismatches 288; Indels 0; Gaps 0;

Qy 814 atacagaagaacaaactgtgggaatttattcttgagaatacgtcacctgatggag 873

Db :

Db 6 www. 65

Qy 874 gatctttactcagcagcggtgcaaaatcgtgcaagatccaaagtatccggggatcac 933

Db :

Db 66 www. 125

Qy 934 tctcaatcatgataatgagtatcttctaccagaagtcagtttgcctgaaagcatta 993

REGISTRATION NUMBER: 34,719
 NAME: Carl J. Schwedler
 REGISTRATION NUMBER: 36,924
 REFERENCE/DOCKET NUMBER: 119-2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (916) 753-6313
 TELEFAX: (916) 753-1510
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1981
 TYPE: nucleic acid
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 US-08-747-574-1

Query Match 40.5%; Score 643.8; DB 3; Length 1981;
 Best Local Similarity 71.8%; Pred. No. 4.1e-167;
 Matches 862; Conservative 0; Mismatches 327; Indels 12; Gaps 1;

QY 317 taagggaattcagatgcaatttccaaagatgtagctttagtggttggggaattg 376
 DB 483 TGAGGAAATACAAAGATTGATCTCAAAAGGGGCATGACTTTGATCTTGAAGCAATG 542

QY 377 gggccaaattgccaacagcaattgtagctatttctcagttgcttcaaaagc 436
 DB 543 GAGACAAATCATACAAATGGCTATGCTATGCTTGGTTCACATTTGTTATCGTTCCAGAGT 602

QY 437 tgatgggttgatgctcctcaagacttgacttgcttactgaaagtgacagtttagact 496
 DB 603 TGATGCCGTGTGATGCTCTTAAACACTTGCTTGTCTTAAAGAAATCGAGATGAGCT 662

QY 497 tgcgaagtgcatttcgaacccctgcattgcagctaatgttgcctgtotccagacttgc 556
 DB 663 TGC AAAATGTATAGCAACCCACTTGTGCGCAAACTGCTGCTGTACAGACTTGCAA 722

QY 557 caatgacactgcaaaacgaatgacagataaaatgtagtattggttggtaaaacagtg 616
 DB 723 CAATGCTGCTGACGAGACCGAATGTCAGATAAATGTTGCTGACTTGTTCGAAAACAGT 782

QY 617 cgtagacagttcaatgagtgagctgctccgaagaatgtagctgtaaatctga 676
 DB 783 GGTGGACCAATTCACGAGTGTGCGGTTTCCGAAAGAAATGTGTGCCCGGAAATCGGA 842

QY 677 tgttggtgacttctgtacctgacacccagtgcttcttccagaaagttgacatgaaaga 736
 DB 843 TGTGGGTGAATTCGCGGTTCGCGATCGTAATGCAGTGGTTCAAAATTTTAAACATGAAAGA 902

QY 737 ttttagcgggaatggttcattactcgcggtttgtaatccacttttgatgtttgattg 796
 DB 903 CTTTAGTGGGAAGTGGTATATAAACAAGTGGTTTAAATCCTACATTTGATGCAATTTGATTG 962

QY 797 ccaattgcatgagttccatacagaagaacaaactgtggggaatttctcttgagaat 856
 DB 963 TCAACTTCATGAGTTTCATATCGAAATGATAAATTTGTTGGAACTTAAATGCGCAT 1022

QY 857 acgtacacctgagaggtattttactcgatcagcgggtgcaaaactctgtgcaagatcc 916
 DB 1023 AAAAATTTGGATGGTGGTCTTTTACTCGATCTGCTGTGCAAACTTTGTTTCAAGATCC 1082

QY 917 aaagtatccgggatactactacaaatcatgataatagtagtatctctcacaagatgactg 976
 DB 1083 AGATCTTCTCGGACACTTATAATCATGACAATGAGTTTCTTCACTACCAAGATGACTG 1142

QY 977 gtatatgttgcatacgaagtgaataatgctccagagattacatttgcataataa 1036
 DB 1143 GTACATATTATCTTCCCAATCGAAACCAACCCGATGATATCATATTCTGATATACCG 1202

QY 1037 gggcagaataatgagtgagtgatgtgtgttctgttactttcacagaagtgagtg 1096
 DB 1203 AGGTCCAAACGACGATGGATGGATACGGTGGGTCCGTGCTGATCTACACCCGAGCCGAC 1262

QY 1097 ttgctggaagcattataccggagttgcaaacccagctcaaaaaagttggcgatatt 1156
 DB 1263 ACTCCCGGAATCGATCATCCCAAACTTACAAAAGCAGCCAAATCCGTGGTGGAGACTT 1322

QY 1157 caacacattcataaaacagacatacatgtgacctgaacctccctctgttggagagtt 1216
 DB 1323 TAACAATTTCAACAAACCCGACAAATAGTTGTTGGCCCTGAGCCTCCATGTTGGTGAAGGCT 1382

QY 1217 ggagaagaagtggagaaggaaggaagcagatcataaaagaagtggagagataagaaga 1276
 DB 1383 TGAGAAAACAGCGGAGAGGGCGAGAGTTGTTGATAAAAAGAGCTGTAGAGATAGAAGA 1442

QY 1277 agaagta-----gagaagtgacagataaagaagtcaccttattcagtaaat 1324
 DB 1443 AGAGTTGAAAAAGAGGTTGAGAGAGGTTAGAGATAGTGTGTTTTCAGAGGTT 1502

QY 1325 gtttgaaggttttaaaagagctcccaagcagatgaagagaacttcttaagagagctgagcaa 1384
 DB 1503 GCTTGAAGGTTTAAAGGAGTTGCAACAAGATGAAGAGAATTTTGTGAGGAGTTGAGTAA 1562

QY 1385 agaagaatggatgttttggatggacttaaaatggagaacactgaggtagaagaaactttt 1444
 DB 1563 AGAAGAGAAGAAATCTGAATGAACTTCAAATGGAAGCGACTGAAGTTGAAAAGCTTTT 1622

QY 1445 tgggcgtgctttaccataaagaaattaaagtaagttatttttaaaactatcaacatatat 1504
 DB 1623 TGGCGCGCGTTACCGATTAGGAACATTAGATAAATTCGATGATTGATTTCAGACAAATAT 1682

QY 1505 a 1505
 DB 1683 A 1683

RESULT 3
 US-08-747-574-3
 Sequence 3, Application US/08747574
 Patent No. 6015939
 GENERAL INFORMATION:
 APPLICANT: CALGENE, INC.
 TITLE OF INVENTION: PLANT VDE GENES AND
 NUMBER OF INVENTION: METHODS RELATED THERETO
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Calgene, Inc.
 STREET: 1920 Fifth Street
 CITY: Davis
 STATE: CA
 COUNTRY: USA
 ZIP: 95616
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB
 COMPUTER: Apple Macintosh
 OPERATING SYSTEM: Macintosh 7.5
 SOFTWARE: Microsoft Word 5.1a
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/747,574
 FILING DATE: No. 6015939 September 7, 1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/023,502
 FILING DATE: August 6, 1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/006,315
 FILING DATE: No. 6015939 September 7, 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Donna E. Scherer
 REGISTRATION NUMBER: 34,719
 NAME: Carl J. Schwedler
 REGISTRATION NUMBER: 36,924
 REFERENCE/DOCKET NUMBER: 119-2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (916) 753-6313
 TELEFAX: (916) 753-1510

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1555
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
US-08-747-574-3

Query Match 39.3%; Score 624.2; DB 3; Length 1555;
Best Local Similarity 74.3%; Pred. No. 8.6e-162;
Matches 820; Conservative 0; Mismatches 268; Indels 15; Gaps 2;

QY 424 ttgcttcaaaagctgacggttgatgcttcaagactgtactgtctactgaaagagt 483
DB 364 TTGTTCCATCTGCAGATCGAGTTGATGACCTTAAACCTTGTCATGCTTATTGAAGGAT 423
QY 484 gcaggttagagcttgcgaagtgcatttcgaacctgcattgcagctaatgtccctgtc 543
DB 424 GCAGGATAGAACTCGCAAGTGCATTCGCAACCTGCTGTGCAGCAATGTCGCTGCC 483
QY 544 tccagactgcaacaataagactgacgaacgaagaaatgcagataaaatgtgtattgt 603
DB 484 TTCAGACCTGCAATTAACCGTCCAGATGAACCGAGTGCAGATTAAATGTGGGATCTGT 543
QY 604 ttgaaaacagtgctgacgagtgatgctgaagtgcagctcccgaaagaaatgtgtac 663
DB 544 TTGACACAGAGTGTGTTGATGAGTTCACGAGTGTGCTGTCTCGAGAAAAAGTGTGTT 603
QY 664 ctgctaaatctgattgtgtgacttctctacatgctccagtgcttctgtccagaagt 723
DB 604 CTAGAAAAATCTGATCTCGGAGAAATTCCTGCCCGCAGACCTTCTGTTCTGTACAGAACT 663
QY 724 ttgacataaagattttagcgggaaatgttctattactcgcggtttgaatccactttg 783
DB 664 TCAACATCTCGGACTTTACGGGAAGTGTGATACATTAAGTGGCTTGAUCCACCTTTG 723
QY 784 atgctttgattgcaattgcatgagttccatcacagaa---gaaacaaacttggggga 840
DB 724 ATGCTTCGACTGCCAGTGCATGAGTGTCCACAGAGAGTGCACAAAGCTTGTGGAA 783
QY 841 atttacttgagaaatcagctacacagagagagatttttactcgatcagcggtgaaa 900
DB 784 ACATCTCTTGAGAAATAAAGACCTTAGACAGTGGATCTTTTACTAGGTGACCCGTACAAA 843
QY 901 aattcgtgcaagatcaaaagtattcggggatactctacaatcatgataatgagtatcttc 960
DB 844 AATTCGTGCAAGATCCTAACCAACCTGTTCTCTACATCATGACAAAGTACCTTC 903
QY 961 tctaccagatgactggttatattttgtcatccaaagttagaataatgctccagagattaca 1020
DB 904 ACTATCAAGATGACTGGTATATCTCTGTCATCAAAAGATAGAGAAATAAACCTGAAGACTATA 963
QY 1021 tatttgtactataaggcgaagaatgcatgagtgatgagtggttctgtacttt 1080
DB 964 TATTTGTATACTACCGTGGCGGAACAGTGTCTGGATGGATATGGTGGCAGCTTGTAT 1023
QY 1081 acacaagaagtgactgtttgctgaaagcattataccggagttgcaaacccgacgtcaaa 1140
DB 1024 ACACGAGAAGTCTGTATTACCCAATAGCATTAACCAAGTACTCGAAAAAGCAGCAAAA 1083
QY 1141 aagttggcggtatttcaaacattcaaaacagacatacatatggtggcctgaaacctc 1200
DB 1084 GCATAGGAGAGACTTCAGCACTTCATTAAGAGGATTAACACATGTGGTCCCTGAACCTG 1143
QY 1201 cccgtttgagaggttggaagaagaagtggaagagagagagacacatcataaagaag 1260
DB 1144 CGCTGTGGAGAGAAATTCAGAGACAGTGGGAAGGTGGAAGGATATATCGTAAAGAGG 1203
QY 1261 ttgagagatagaagaagaatgagagaggtgagagataaagaag-----tca 1308
DB 1204 TTGAAGAGATAGAAGAGAGGTAGAGAAGGAAGTGGAGAAGGTCGGTAGGACTGAGATGA 1263

QY 1309 ccttattcagtaaaactgtttgaaggttttaaagagctccaacgagatgaagagaactttc 1368
DB 1264 CCTGTTCCAGAGATTGGCTGAAGGATTTAAATGAACCTGAAGCAAGACGAGGAGATTTCG 1323
QY 1369 taagagagctgagcaagaagaatggtatgttttggatggacttaaaatggaagcaactg 1428
DB 1324 TGAGAGAGTTAAGTAAGAAGAGATGGAGTGTGATGAGATCAAAATGGAAGCAAGTG 1383
QY 1429 agttagaaaaacttttggcgctgttaccataaagaaaataaggttaagtatttttaa 1488
DB 1384 AGTTGAAAAATTTGTTGGGAAAGCTTTCCTCAATCAGGAAGGTTCAGGTAGTAACCAAGAAC 1443
QY 1489 aactatcaacatatatactacat 1511
DB 1444 CACCATGTTGTACAAACTATAT 1466

RESULT 4

US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 4.8%; Score 75.8; DB 1; Length 7218;

Best Local Similarity 5.6%; Pred. No. 4.3e-11;
Matches 23; Conservative 237; Mismatches 149; Indels 0; Gaps 0;

QY 999 gaaatagtcagagagattacatatattgttactataagggcagaaatgatcgtggat 1058

. Db 3221 AGGAAGAGAGGAGGAAGAAGACTTTGAGGAAGAATTTGAGGAAGAAGAGGTGAGTTAG 3280
 . QY 1435 aaaaa 1439
 . Db 3281 AGGAA 3285

```

RESULT      9
US-08-973-462-2
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5361)
US-08-973-462-2

```

[illegible]

```

RESULT 10
US-08-973-462-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-BRYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12

```

```

; EARLIER APPLICATION NUMBER: FR 95/070007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
US-08-973-462-1

```

```

Query Match      3.1%; Score 48.6; DB 4; Length 6152;
Best Local Similarity 47.8%; Pred. NO. 0.0011;
Matches 141; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 1221 agaaaagtggaaagggagaaagacgatcataaaaaagttgaggagatagaagaagaa 1280
    || || || || || || || || || || || || || || || || || || || || ||
Db 4163 aaagatgtcgaaagagacaagatcgaaaagtattctgattaaagatcttgaagaagat 4222
    || || || || || || || || || || || || || || || || || || || || ||
QY 1281 gtagaagaaggtgagagataaagaagtcaccttattcgataaacctgtttgaaggttttaa 1340
    || || || || || || || || || || || || || || || || || || || || ||
Db 4223 atattaaaagaagtgaaaagagaatcaaaagaaacttgaaagtgaaattttagaagattataa 4282
    || || || || || || || || || || || || || || || || || || || || ||
QY 1341 gagctccaaacgagatgaagagaactctttaagagagctgagcaagaagaatacggatgtt 1400
    || || || || || || || || || || || || || || || || || || || || ||
Db 4283 gaattaaaactattgaaacagagatttttagaagagaaaaagaatacagaaaagaatcat 4342
    || || || || || || || || || || || || || || || || || || || || ||
QY 1401 ttggatggacttaaaatggaagaaactgaggtagaaaaaactttttggggtgctttacca 1460
    || || || || || || || || || || || || || || || || || || || || ||
Db 4343 ttgtaaaaattcgaaagaagacgtaagaataaaagaatctctgaagcagatatattaaaaa 4402
    || || || || || || || || || || || || || || || || || || || || ||
QY 1461 ataaggaaaattaaagtgtaagtatttttttaaaactatcaacatatatactacatgtat 1515
    || || || || || || || || || || || || || || || || || || || || ||
Db 4403 qaagatctcttcattagaacttgaagaacaaaaaaaatttagaacaagttacacaaat 4457
    || || || || || || || || || || || || || || || || || || || || ||

```

```

REMARK      11
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

```

```

Query Match      2.8%; Score 45.2; DB 4; Length 289;
Best Local Similarity 9.3%; Pred. No. 0.0023;
Matches 25; Conservative 111; Mismatches 134; Indels 0; Gaps 0;
QY 1209 gagaggttgagaaacuaaagtgtgaagaaaggaagacgcgatcataaaaagattgaagaa 1268

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2001, 00:54:03 ; Search time 3841.92 Seconds
(without alignments)
3909.656 Million cell updates/sec

Title: US-09-075-375A-3
Perfect score: 1589
Sequence: 1 tattttcatgagtttgcagt.....tctttttcagaagcaaaaaa 1589

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST : *

1:	gb_est1.*	72:	em_estp11.*
2:	gb_est2.*	73:	em_estp12.*
3:	gb_est3.*	74:	em_estp13.*
4:	gb_est4.*	75:	em_estp14.*
5:	gb_est5.*	76:	em_estp15.*
6:	gb_est6.*	77:	em_estp16.*
7:	gb_est7.*	78:	em_estp17.*
8:	gb_est8.*	79:	em_estp18.*
9:	gb_est9.*	80:	em_estp19.*
10:	gb_est10.*	81:	em_estp110.*
11:	gb_est11.*	82:	em_estro1.*
12:	gb_est12.*	83:	em_estro2.*
13:	gb_est13.*	84:	em_estro3.*
14:	gb_est14.*	85:	em_estro4.*
15:	gb_est15.*	86:	em_estro5.*
16:	gb_est16.*	87:	em_estro6.*
17:	gb_est17.*	88:	em_estro7.*
18:	gb_est18.*	89:	em_estro8.*
19:	gb_est19.*	90:	em_estro9.*
20:	gb_est20.*	91:	em_estro10.*
21:	gb_est21.*	92:	em_estro11.*
22:	gb_est22.*	93:	em_estro12.*
23:	gb_est23.*	94:	em_estro13.*
24:	gb_est24.*	95:	em_estro14.*
25:	gb_est33.*	96:	em_estro15.*
26:	gb_est34.*	97:	em_estro16.*
27:	gb_est35.*	98:	em_estro17.*
28:	gb_est36.*	99:	em_estro18.*
29:	gb_est37.*	100:	em_estro19.*
30:	gb_est38.*	101:	em_estro20.*
31:	gb_est39.*	102:	gb_est25.*
32:	gb_est40.*	103:	gb_est26.*
33:	em_estba.*	104:	gb_est27.*
34:	em_estfun.*	105:	gb_est28.*
35:	em_esthum1.*	106:	gb_est29.*
36:	em_esthum2.*	107:	gb_est30.*
37:	em_esthum3.*	108:	gb_est31.*
38:	em_esthum4.*	109:	gb_est32.*
39:	em_esthum5.*	110:	gb_est41.*
40:	em_esthum6.*	111:	gb_est42.*
41:	em_esthum7.*	112:	gb_est43.*
42:	em_esthum8.*	113:	gb_est44.*
43:	em_esthum9.*	114:	gb_est45.*
		115:	gb_est46.*
		116:	gb_est47.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est167:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*

190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.


```

Db 568 GTCTCCAAACATGCAACAATAGACCTGATGAACGGAATGTCAGATAAAATGTGGTGATT 627
QY 601 tgtttgaaacagtgtcgta 620
Db 628 TGTTGAAACAGTGTGTGA 647

RESULT 2
LOCUS BG123331 724 bp mRNA EST 31-JAN-2001
DEFINITION EST468977 tomato shoot/meristem Lycopersicon esculentum cDNA clone
cTOFLJ5 5' sequence, mRNA sequence.
ACCESSION BG123331
VERSION BG123331.1 GI:12623519
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 724)
AUTHORS van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Utterback,T.,
Hansen,C., Ronning,C. and Tanksley,S.
TITLE Generation of ESTs from tomato shoot/meristem tissue
JOURNAL Unpublished (2001)
COMMENT Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.

FEATURES
source
1..724
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOPLJ5"
/clone_lib="tomato shoot/meristem"
/tissue_type="shoot/meristem"
/dev_stage="developing shoots from 4-6wks old plants"
/lab_host="SOLR"
Xhol: Small expanding leaves from the growing tip were
taken from greenhouse plants (4-6wks old TA496). Tissue
was immediately frozen in liquid nitrogen."
BASE COUNT 221 a 131 c 158 g 214 t
ORIGIN

Query Match 31.2%; Score 495.4; DB 174; Length 724;
Best Local Similarity 85.1%; Pred. No. 1.9e-117;
Matches 578; Conservative 0; Mismatches 71; Indels 30; Gaps 1;

QY 1 tattttcatgagtttcagtttggtggaatacaggttggaagaatggtcttgcctccatcatt 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 45 TATTATCATGAGTTGGAGTTGGTGTGCTATTCATTAATGAATAATGCGCTTGCCCTCAT 104
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 caaatcttcggccaaccatgaacacataatattatgttggtcaaaagctcccggtc 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105 CAACAATCTGTGCAATCATGAGGCCATCAATGTCAGATTTGGATCGAGCTTCAGAGTC 164
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 ataaaggtttagctggggttggaagattacttttggtagtagtagctagcagaataatt 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165 ATACAAGATTTAGCTGGGTAGAGCAGAGTACTTTGGTAGTAGTATGCTAGTGAATAATT 224
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 gttccagagacagatacactagatactttcgaaatctctctagaatagctgtggttgg 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 225 GTTCCAGAGACAGATACATACATCTTCAGAAAATCTTCTAGAATATGTTGTGTTGG 284
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 attcaagagttctgcaactattctcacacggggaacacacaaatctctctcccgacacatgca 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 285 ATTCTAGAGTCTGCAACTATCATCATCGGGGGGAACAAATCTTCTCTCGACATAGAA 344
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY 301 ttaaccagaatgtacctaagggaatcagagatgcaaatctccaaagatgtagctttga 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 345 TTACCAGAAATGTACTTAAGGAAATACAATATGGAATTTCCAGAAATGATGCTTTGA 404
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 tggttgggagaatggggccaatttccaaacacagcaattgttagctatatattttgt 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 405 TGGTCTTGAGAAATGGGGCAATTTGGCCAAACACAGCAATTTGTAACATATATTTTGT 464
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 cagttgttcaaaagctgtagcggttgatgctctcaagactgttactgttactgaaag 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 465 CAGTTGCTTTCAAAGGCTGATGCCGTTGATGCTCTCAAAACTGTACTTGTCTACTGAAAG 524
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 agtgcaggt-----tagagcttgcgaagtgcattt 510
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 525 AGTCAGGTAATTTTCAGATGTTACTTAAGTATGCGCGATAGAGCTTGCGAAGTGCATCT 584
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 511 cgaacctgcattgcagctaaattgttcctgtctccagacttgcacaatatagacctgacg 570
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 585 CAACCCCTGCAATGTCAGCTAATGTCCTCTGCCAAACATGCAACAATAGACCTGATG 644
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 571 aaacggaatgcagataaaatgtggtgatttggttgaaacagtgctgtagacgagttca 630
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 645 AAACGGATGTCAGATAAAATGCTGCTGATTTGTTGAAACAGTGTGTAGACAGTTCA 704
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 631 atgagttgcagctctcccg 649
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 705 ATGAGTGTGCAGTCTCCCG 723
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
LOCUS AI483033 584 bp mRNA EST 29-JUN-1999
DEFINITION EST242356 tomato shoot, Cornell Lycopersicon esculentum cDNA clone
CLEB8K10, mRNA sequence.
ACCESSION AI483033
VERSION AI483033.1 GI:4386957
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 584)
AUTHORS van der Hoeven,R.S., Matern,A.L., Vision,T., Holt,I.E., Liang,F.,
Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L.,
Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Giovannoni,J.J.
and Tanksley,S.D.
TITLE Generation of ESTs from tomato shoot meristem
JOURNAL Unpublished (1999)
COMMENT Contact: David Frisch
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@clemson.edu.
FEATURES
source
1..584
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEB8K10"
/clone_lib="tomato shoot, Cornell"
/tissue_type="shoot meristem"
/dev_stage="8 week old plants"
/lab_host="XIOUR"
Xhol: Vector: pBK_CMV; Site_1: EcoRI; Site_2: XhoI; cLEB
- Tomato Shoot Meristem EST Library. Oligo-dT primed cDNA
library made from tomato vegetative shoots including
meristems and small expanding leaves."
BASE COUNT 179 a 110 c 124 g 171 t
ORIGIN

```



```

Query Match      26.6%; Score 423.2; DB 20; Length 584;
Best Local Similarity 87.2%; Pred. No. 8.6e-99;
Matches 464; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1 tattttcatgagtttgcaagtttggtgtaataacggttgagaagatggctcttggccctcatt 60
Db 53 TATTTACATGAGTGTGGAGTGTGGTGTCTCAATTAATTAATTAATTAATTAATTAATTAAT 112
QY 61 caaatcttctggcccaacatgaacacataatattatgttgggtcacaagcttcccggtc 120
Db 113 CAAACATCTGTGCAATCATGAGGCCATCAATGTCAGTTGGATCAAGGCTTCAGAGTC 172
QY 121 ataaagagtttgagctgggttggaagattactttggttagtatagctcagcaaaattt 180
Db 173 ATACAGATTATTAGCTGGGGTAGAGCAGATTACTTTGGTAGTATAGTCTAGTGAAATTT 232
QY 181 gttccagcagcaggtaccatagatacttctgaaaatctcctagaatagctggtgttg 240
Db 233 GTTCCAGACAGATACCTACATCTGACAGAAATCTTCTAGAATATGTTGGTGTGG 292
QY 241 attcaagaggtctcaactattctcacacggggaacacaaatctctcccgacacatagca 300
Db 293 ATTCTAGAAGTCTGCAACTATCATCAGGGGGGAAACAAATCTTCTTCGCACATAGAA 352
QY 301 ttaaccagaatttacctaaagggaatttcaggatgcaaatcttccaaaagatgtagcttga 360
Db 353 TTAAACAGAAATGTACTTAAGGGAATACAATATGGAATTTCCAGAAAGATGATAGCTTTGA 412
QY 361 tgggttggagaaatggggcccaattgccaacacagcaattgtagctatattctttgt 420
Db 413 TGGTCTTGAAGAAATGGGGCCCAATTTGGCCAAACAGCAATGTACTATATTTATTTGT 472
QY 421 cagttgcttcaaaagctgatcggttgatgctctcaagaacttgtagcttactgctactgaaag 480
Db 473 CAGTTGCTTCAAAAGGCTGATGCGGTGATGCTCTCAAAACTTGTACTTGTACTGTAAG 532
QY 481 agtgcaggttagagcttgcaagtagcatttcgaacccctgcagtgcactaa 532
Db 533 AGTGCAGGATAGAGCTTGGGAAGTGCATCTCAAAACCCCTGCATGTGCAGCTAA 584

RESULT 4
AW929268 572 bp mRNA EST 30-MAY-2000
LOCUS EST338056 tomato flower buds 8 mm to pre-anthesis, Cornell
DEFINITION University Lycopersicon esculentum cDNA clone cT0C715 5', mRNA
sequence.
ACCESSION AW929268
VERSION AW929268.1 GI:8104669
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 572)
AUTHORS van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang
,F., Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
,C.M., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
Generation of ESTs from tomato flower tissue
Unpublished (1999)
Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU
5 prime sequence.
LOCATION/Qualifiers

FEATURES
SOURCE source
1..572
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cT0C715"
/clone_lib="tomato flower buds 8 mm to pre-anthesis,
Cornell University"
/tissue_type="flower"
/dev_stage="buds 8mm-to-preanthesis"
/notes="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Tanksley; Flower buds and flowers were
taken from greenhouse plants (4-8 wks old, TA496). They
were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."
BASE COUNT 175 a 107 c 120 g 170 t
ORIGIN

Query Match      26.3%; Score 418.2; DB 122; Length 572;
Best Local Similarity 87.1%; Pred. No. 1.7e-97;
Matches 459; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1 tattttcatgagtttgagttgggttaataacggttgagaagatggctcttggccctcatt 60
Db 46 TATTTACATGAGGTTGGAGTTGGTGTCAATTAATTAATTAATTAATTAATTAATTAAT 105
QY 61 caaatcttctggcccaacatgaacacataatattatgttgggtcacaagcttcccggtc 120
Db 106 CAAACATCTGTGCAATCATGAGGCCATCAATGTCAGTTGGATCAAGGCTTCAGAGTC 165
QY 121 ataaagagtttagctgggttggaagattactttgtagtatagctcagcaaaattt 180
Db 166 ATACAAGATTTAGCTGGGGTAGAGCAGATTACTTTGGTAGTATAGTCTAGTGAAATTT 225
QY 181 gttccagcagcaggtaccatagatacttctgaaaatctcctagaatagctggtgttg 240
Db 226 GTTCCAGACAGACATACCTACATCTTGCAGAAATCTTCTAGAATATGTTGGTGTGG 285
QY 241 attcaagaggtctcgaactattctcacacggggaacacaaatctctcccgacacatagca 300
Db 286 ATTCTAGAAGTCTGCAACTATCATCAGGGGGGAAACAAATCTTCTTCGCACATAGAA 345
QY 301 ttaaccagaatttacctaaagggaatttcaggatgcaaatcttccaaaagatgtagcttga 360
Db 346 TTAAACAGAAATGTACTTAAGGGAATACAATATGGAATTTCCAGAAAGATGATAGCTTTGA 405
QY 361 tgggttggagaaatggggcccaatttggccaaaacagcaattgtagctatattctttgt 420
Db 406 TGGTCTTGAAGAAATGGGGCCCAATTTGGCCAAACAGCAATGTACTATATTTATTTGT 465
QY 421 cagttgcttcaaaagctgatcggttgatgctctcaagaacttgtagcttactgctactgaaag 480
Db 466 CAGTTGCTTCAAAAGGCTGATGCGGTGATGCTCTCAAAACTTGTACTTGTACTGTAAG 525
QY 481 agtgcaggttagagcttgcaagtagcatttcgaacccctgcagtgcga 527
Db 526 ACTGCAGGATAGAGCTTGGGAAGTGCATCTCAAAACCCCTGCATGTGCA 572

RESULT 5
BG096454 444 bp mRNA EST 29-JAN-2001
LOCUS EST460973 potato leaves and petioles Solanum tuberosum cDNA clone
DEFINITION CSTRB43H21 5' sequence, mRNA sequence.
ACCESSION BG096454
VERSION BG096454.1 GI:12586489
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 444)

```


242	CTGTGGAAAGCGCTGGAGAAAAGGTAGAGAAAGGAGAGACCAATTGTTAAGGGAAGT	301
1262	tgaggagatagaagaagaagtgaagaagtgaagataaaagaagtcacacttattcagtaa	1321
302	TGAACAGTTTAGAAGAAGAGGCTGGAGAAAGTGGGGGAAAACGGAGGCCACCTTGTTCAAAA	361
1322	actgtttgaaggttttaagaagctccacgcgatatgaagaactcttaagagagctgag	1381
362	ATTTCGAGAAAAGGTTCAAAAGTATTCAAAAGAAATGAGAGAAATTCCTTAAGAAAGTTATC	421
1382	caagaagaatacgatggttttggatggacctaaatggaagcaactgaggttagaaaaact	1441
422	AAAAACAAGAAATGGAATAATTCCTTATAGGCTCAAAATGGAACCCAGTCAGTAAATAAAAAACT	481
1442	ttttggcgctgctttaccataaggaataaag	1474
482	CTTTAGACGTGCGCTTCGCGTGGAAGAAACCTAAG	514

RESULT 7

AI994713
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

483 bp mRNA
EST
08-SEP-1999
cDNA clone 701499353 A. thaliana, Ohio State clone set Arabidopsis thaliana
cDNA clone 701499353, mRNA sequence.
AI994713.1 GI:5841618
EST.
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons: core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 483)
Chen, J., Momtazi, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,
Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P.,
Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R.,
Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T.,
Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobrega, A., Murry, L.,
Turner, C., Krikorian, S., Elder, L. and Hanson, D.
Arabidopsis thaliana Gene Expression MicroArray
Unpublished (1999)
Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.

606	gaaacagtgctgtagacgaggttcaatgagtgtagtcgagtcctccgcgaagaaatggtgtacct	665
134	GAGAACAGTGTGTTGATGAGTCAACGAGTGTGCTGTGTCGAGAAAAAGTGTGTTCT	193
666	cgtaaatctgattgttggaacttctctgacctgacccagtggtctcttcaccagaagttt	725
194	AGAAATCTGATCTCGGAGAAATTTCTGCCCCAGACCCTTCTGTTCTGTACAGAACTTC	253
726	gacatgaagatttttagcggggaattggttcattactcgcggtttgaatccccacttttgat	785
254	AACATCTCGGACTTTAAACGGGAAGTGGTACATTACAAGTGGCTTGAATCCAACCTTTGAT	313
786	gcttttgattgccaattgcataagttccatacagaa---gaaacaaaccttgggggaat	842
314	GCTTTCCGACTGCCAGTGCATGAGTTCACAGAAAGGTGACAAAGCTTGTGTGGAAC	373
843	ttatcttgagaatacgtacacactgatggaggtatttttactcgcatacgcggtgcacaaa	902
374	ATCTCTGGAGATAAAGACCCTAGACAGTGGATCTTTTACTAGGTCACGCCGTACAAAA	433
903	ttcgtgcaagatccaaagtatccggggatactctacaatcatgataatga	952
434	TTTCGTGCAAGATCCTCACCAAACTGGTGGTCTCTACAATCATGACAACGA	483

RESULT	8	
AV410579		
LOCUS	427 bp mRNA EST 23-MAY-2000	
DEFINITION	AV410579 Lotus japonicus young plants (two-week old) Lotus japonicus cDNA clone MW1074g05_r 5', mRNA sequence.	
ACCESSION	AV410579	
VERSION	AV410579.1 GI:7723433	
KEYWORDS	EST.	
SOURCE	Lotus japonicus.	
ORGANISM	Lotus japonicus	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotus.	
AUTHORS	Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.	
TITLE	Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus	
JOURNAL	Journal of Plant Gene Research	
MEDLINE	DNA Res. 7 (2), 127-130 (2000)	
COMMENT	Contact: Yasukazu Nakamura The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: ynakamu@kazusa.or.jp; URL: http://www.kazusa.or.jp/en/plant/. Location/Qualifiers 1. .427 /organism="Lotus japonicus" /db_xref="taxon:34305" /clone="MW1074g05_r" /clone_lib="Lotus japonicus young plants (two-week old)" /dev_stage="young plants (two-week old)" /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; isolate=Miyakojima MG-20"	
BASE COUNT	112 a 85 c 102 g 128 t	
ORIGIN		
Query Match	14.6%; Score 232.2; DB 30; Length 427;	
Best Local Similarity	76.4%; Pred. No. 2.1e-49;	
Matches 285; Conservative	0; Mismatches 88; Indels 0; Gaps 0;	
QY	409 tattcattttgcagttgcttcaaaagctgacgcggttgatgctctcaagacttgctactt	468
Db	55 TGTGCATTTTAATGATCGTCCACAGCTGATGCTGTTGATGCTCTCAAAACTTGTCTT	114
QY	469 gcttactgaaagtgtaggttagcttgcgaagtgcatcttcgaaccttgcatgtgcag	528
Db	115 GTTGTGCTGAAGTAATCAGGATTGAATGGCTGAAGTGATCTCAAAACCGCTTTGTGCTG	174

/strain="Columbia"
/db_xref="taxon:3702"
/clone="T27M8"
/clone_lib="TAMU"
/sex="hermaphrodite"
/note="Vector: BelobACII; Site_1: HindIII; Site_2: HindIII
; Produced by Rod Wing"
BASE COUNT 204 a 365 c 195 g 357 t 25 others
ORIGIN

Query Match 9.9%; Score 156.6; DB 256; Length 1147;
Best Local Similarity 76.2%; Pred. No. 9.7e-30;
Matches 192; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 578 atgtcagataaaatgtgtgtattgtttgaaacagtgctagacaggttcaatgagt 637
DB 324 ATATCATGATTAAATGGGGATCTCTTTGAGAACAGTGTCTTGTATGAGTTCAACGAGTG 265
QY 638 tgcagctcccgaaagaaatgtacctcgtaaatctgattgtgtgactttcctgtacc 697
DB 264 TGCTGTGCGAGAAAAGTGTGTTCTAGAAAATCTGATCTCGAGAAATTTCTGCCCC 205
QY 698 tgatcccgagttctgtccgaaagtgtgacatgaaagatttttagcggaagtgttcat 757
DB 204 AGACCTCTTCTGTTCTGTACAGAACTTCAACATCTCGGACTTTAAACGGGAAGTGTACAT 145
QY 758 tactcgcggttgatcccaacttttgatgctgtttgattgccaatgcatgagttccatcac 817
DB 144 TACAAGTGGCTTGAATCAACCTTTTTCCTTCGANTGCCAGCCCCATGATGTTCCACAC 85
QY 818 aqaagaaacaa 829
DB 84 AGAAGTGACAA 73

RESULT 15
A2126249/c
LOCUS A2126249 812 bp DNA GSS 02-JUN-2000
DEFINITION OSUNBB0075L17f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
clone OSUNBB0075L17f, DNA sequence.

ACCESSION A2126249
VERSION A2126249.1 GI:8200598
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 812)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
Contact: Wing RA
Clemson University
Clemson University Genomics Institute

100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7286
Fax: 864 656 4293
Email: rwing@clemson.edu

Seq primer: GTAAAACGACGGCCAGTG
Class: BAC ends
High quality sequence start: 40
High quality sequence stop: 642.

Location/Qualifiers
1..812
/organism="Oryza sativa"
/strain="Japanica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="OSUNBB0075L17f"
/clone_lib="CUGI Rice BAC Library (EcoRI)"
/tissue_type="Leaf"

FEATURES
source

/lab_host="E. coli DH10B"
/note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa,
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 Kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation of a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT 234 a 161 c 182 g 227 t 8 others
ORIGIN

Query Match 9.8%; Score 156.2; DB 238; Length 812;
Best Local Similarity 69.3%; Pred. No. 1.1e-29;
Matches 212; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 578 atgfcagataaaatgtgtgtattgtttgaaacagtgctgtagacgagttcaatgagt 637
DB 320 ATTCTAGATCAATGTGGAGATCTGTTGAGAACACTGTGTGTCGATCAGTTCAACGAGTG 261
QY 638 tgcagctcccgaaagaaatgtgtacctgtgtaaatctgtatgttggtagacttctctgacc 697
DB 260 TGCTGTTTCGCGCAAGAAATCGCTCCACAAAAGTCCGAGTTGGCGAGTTCACGATCCC 201
QY 698 tgatcccgagttctgtccgaagtgtgacatgaaagatttttagcggaagtgttcat 757
DB 200 TGATCCATCCGCCCTTCTCAAGAACTTCAACATGGCTGATTTCAACGCAAGTGGTATAT 141
QY 758 tactcgcggtttgaatccccacttttgatgcttttgatgccaattgcatgagttccatcac 817
DB 140 TTCAAGTGGCTCAATCCCACCTTCGACACATTCGATTCGCACTTCACGAGTTCCGTGT 81
QY 818 agaagaaacaaacttgtggggaatttatcttggagaaatcacgtacacctgatggagatt 877
DB 80 CGAGGGAGACAAACTTATAGCAACTTGACATAGAGAAATTCGGCGCTATAGACGGA 21
QY 878 tttta 882
DB 20 TCTAA 16

Search completed: November 6, 2001, 00:54:09
Job time: 4098 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 04:59:28 ; Search time 93.61 Seconds
(without alignments)
914.444 Million cell updates/sec

Title: US-09-075-375A-4
Perfect score: 7495
Sequence: 1 MALSHTVFLCKEALNYA.....MEASEVEKLFKALPIRKVR 1412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2554	34.1	478	18 AAW09875	Tobacco violaxanthin
2	2498	33.3	473	18 AAW09874	Romaine lettuce vi
3	2417	32.2	462	18 AAW09876	Arabidopsis violax
4	212.5	2.8	525	21 AAG49996	Arabidopsis thalia
5	210	2.8	522	21 AAG49997	Arabidopsis thalia
6	209	2.8	522	21 AAG06331	Arabidopsis thalia
7	209	2.8	526	21 AAG06330	Arabidopsis thalia
8	207	2.8	428	21 AAG49998	Arabidopsis thalia
9	206	2.7	428	21 AAG06332	Arabidopsis thalia
10	151	2.0	1055	21 AAY44787	Arabidopsis thalia
11	147.5	2.0	1786	18 AAW24790	P. falciparum live

12	147	2.0	2482	16 AAR72826	Human mitotin. Ho
13	147	2.0	2482	19 AAW23996	Human mitotin amin
14	143.5	1.9	1558	21 AAB18324	Plasmodium falcipa
15	135.5	1.8	1312	18 AAW22775	Human RAD50. Homo
16	135.5	1.8	1312	19 AAW71295	Human homologue of
17	135	1.8	968	19 AAW85011	p85alpha-green flo
18	134	1.8	1639	19 AAW54145	P. falciparum synt
19	133	1.8	1979	21 AAB18171	Plasmodium falcipa
20	132	1.8	776	15 AAR60178	Lethal factor of B
21	129.5	1.7	2954	20 AAY01632	Amino acid sequenc
22	129	1.7	993	17 AAY95268	Pre-nisin modifica
23	129	1.7	993	20 AAY06666	Nisin B of lactoba
24	129	1.7	1654	6 AAP50777	Sequence of the P1
25	128.5	1.7	1521	21 AAG39235	Arabidopsis thalia
26	128.5	1.7	1528	21 AAG39234	Arabidopsis thalia
27	128.5	1.7	1562	21 AAG39233	Arabidopsis thalia
28	128.5	1.7	1703	21 AAG36714	Arabidopsis thalia
29	128.5	1.7	1710	21 AAG36713	Arabidopsis thalia
30	128.5	1.7	1744	21 AAG36712	Arabidopsis thalia
31	124	1.7	1498	16 AAR77084	Rat sulphonylurea
32	124	1.7	1582	16 AAR77087	Rat sulphonylurea
33	121.5	1.6	1132	17 AAR97866	Chicken leucocytos
34	121.5	1.6	1369	20 AAY24788	Human secreted pro
35	121	1.6	1411	17 AAW02258	Nucleolar/endosoma
36	119.5	1.6	1099	19 AAW40538	Mutant C-beta prot
37	119.5	1.6	1604	16 AAR70105	TNF-R-EBA 175 fusi
38	119.5	1.6	1847	21 AAY52002	M. jannaschii MJ14
39	119.5	1.6	1847	21 AAY51631	M. jannaschii MJ14
40	119.5	1.6	1881	21 AAY44506	Streptococcus pneu
41	119	1.6	905	18 AAW31186	Human p160 polyep
42	119	1.6	1135	18 AAW31185	Human p160 polyep
43	118.5	1.6	2273	17 AAR98811	Erysiphe graminis
44	117.5	1.6	1230	21 AAB49306	Potato starch synt
45	117.5	1.6	3542	22 AAB62142	P. falciparum PCR3

ALIGNMENTS

RESULT 1
ID AAW09875 standard; Protein; 478 AA.
XX
AC AAW09875;
XX
DT 28-JUL-1997 (first entry)
XX
DE Tobacco violaxanthin de-epoxidase.
XX
KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
KW xanthophyll; tobacco.
XX
OS Nicotiana tabacum cv. xanthi.
XX
FH Key Location/Qualifiers
FT Peptide 1..134
FT /label= Transit_peptide 135..478
FT Protein /label= Mat_protein 135..147
FT Peptide /note= "Claim 8"
FT Domain 135..206
FT /label= Cys-rich_domain 385..478
FT Domain /label= Highly-charged_domain 141
FT Misc-difference /note= "conserved Cys residue"
FT Misc-difference 143
FT /note= "conserved Cys residue"
FT Misc-difference 148
FT /note= "conserved Cys residue"
FT Misc-difference 155

FT Misc-difference 161 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 167 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 171 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 180 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 184 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 190 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 206 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 252 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 FT Misc-difference 382 /note= "conserved Cys residue"
 FT /note= "conserved Cys residue"
 XX
 PN WO9717447-A2.
 XX
 XX 15-MAY-1997.
 XX
 PD 07-NOV-1996; 96WO-US18291.
 XX
 PF 06-AUG-1996; 96US-00233502.
 XX
 PR 07-NOV-1995; 95US-0006315.
 XX
 XX (CALJ) CALGENE INC.
 PA
 XX
 XX Bugos RC, Rockholm DC, Yamamoto HY;
 PI
 XX
 DR WPI; 1997-281036/25.
 DR N-PSDB; AAT66242.
 XX
 PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light
 XX
 PS Disclosure; Fig 2; 4lpp; English.
 XX
 CC The 55 kDa violaxanthin de-epoxidase (VDE) (AAW09875) of tobacco
 CC catalyses the de-epoxidation of violaxanthin to zeaxanthin and
 CC antheraxanthin. This system, termed energy dependent non-radiative
 CC energy dissipation or non-photochemical fluorescence quenching,
 CC reduces the quantum efficiency of photosystem II (PSII), helping to
 CC prevent PSII over-reduction and photoinhibitory damage. The amino
 CC acid sequence of the VDE was deduced from an isolated cDNA clone
 CC (AAT66242). VDE nucleic acids (see also AAT66241, AAT66243), in sense
 CC or antisense orientation, can be used in genetic constructs to
 CC modify VDE levels in plants. Increased levels result in the plant
 CC being tolerant of increased light and therefore more productive
 CC and/or more resistant to disease. Underexpression of VDE increases
 CC photosynthetic efficiency under low light. The photosensitivity of
 CC a range of crops, trees and ornamentals can be modified.
 XX
 SQ Sequence 478 AA;

Query Match 34.1%; Score 2554; DB 18; Length 478;
 Best Local Similarity 99.8%; Pred. No. 3.9e-191;
 Matches 476; Conservative 1; Mismatches 0; Gaps 0;

Qy 474 MALAPHSNFIANHETIKYYVGSKLPGHKRFSWGWEYFGSIVVAKICSSRRIPRYRKSP 533
 Db 1 malaphsnflanhetikyygsklpghkrfswgweyfgsivvakiessrripryrksp 60
 Qy 534 RICGDLRSGLQFSGHKNLSPAHSINQNPVKGNSGCKFPKDVAlmVWKGQFARTAI 593
 Db 61 riccgldrsqvgfshgkhnlsphaisinqnpvkgngskfpkdvalmVWKGQFARTAI 120
 Qy 594 VAIFILSVASKADAVDAKTCCLLKECRLELAKCISNPACAAVACLOTCNNRPDETCC 653
 Vaifilsvaskadavdalktctclllkecrlelakisnpacaanvacloctcnnrpdetcc 180

Db 121 vaifilsvaskadavdalktctclllkecrlelakisnpacaanvacloctcnnrpdetcc 180
 Qy 654 QIKCGDLFENSVDENECASVRKKCVPRKSDVGDPPVDPSPVLQKFDKMDFGKWFIT 713
 Db 181 qikcgdlfensvvdedefnecavsrkkcvprksdvgdppvdpvslvqkfdmkdfsgkwfit 240
 Qy 714 RGLNPTFDADFDCOLHEFHTEENKLVGNLSWRIRTPDGGFEFTRSAVOKFVQDPKYPGILYN 773
 Db 241 rglnpftfdadfcolhefhteenkvlgnlswrirtpdggfftrsvavqkfvgdpkypgillyn 300
 Qy 774 HDNEYLLYQDDWYLSSKVENSPEDYIFVYKGRNDWDGYSGLYTRSAVLPEIPIE 833
 Db 301 hdneyllyqddwylsskvenspedyifvykgrndwdgysgsvlytrsavlpesilpe 360
 Qy 834 LQTAAGKVGGRDFNTFIKTNTCGPEPLVERLEKKVVEGERTIIKEVEIEVEKVRDK 893
 Db 361 lqtaagkvgrdfntfiktntcgppeplverlekkvveegertiikeveeleeevekevrdk 420
 Qy 894 EVTLFSLKLFEGFKELQDENFLRELKSEMDVLDGLKMEATEVEKLFGRALPIRLK 950
 Db 421 evtlfslklfegfkqlqrdeenflrelskeemdvldglkmeateveklfgralpirkl 477

RESULT 2
 AAW09874
 ID AAW09874 standard; Protein; 473 AA.
 AC AAW09874;
 XX
 DT 28-JUL-1997 (first entry)
 XX
 DE Romaine lettuce violaxanthin de-epoxidase.
 XX
 KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 KW xanthophyll; lettuce.
 XX
 OS Lactuca sativa L. cv. romaine.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..125
 FT /label= Transit_peptide
 FT Protein 126..473
 FT /label= Mat_protein
 FT Peptide 126..138
 FT /note= "Claim 8"
 FT Domain 126..197
 FT /label= Cys-rich_domain
 FT Peptide 218..231
 FT /label= Lipocalin_signature
 FT Domain 376..473
 FT /label= Highly-charged_domain
 FT Peptide 265..272
 FT /label= Tryptic_peptide-11
 FT Peptide 275..289
 FT /label= Tryptic_peptide-21
 FT Peptide 341..353
 FT /label= Tryptic_peptide-15
 FT Misc-difference 132 /note= "conserved Cys residue"
 FT Misc-difference 134 /note= "conserved Cys residue"
 FT Misc-difference 139 /note= "conserved Cys residue"
 FT Misc-difference 146 /note= "conserved Cys residue"
 FT Misc-difference 152 /note= "conserved Cys residue"
 FT Misc-difference 158 /note= "conserved Cys residue"
 FT Misc-difference 162 /note= "conserved Cys residue"
 FT Misc-difference 171 /note= "conserved Cys residue"


```
XX 07-NOV-1996; 96WO-US18291.
XX 06-AUG-1996; 96US-0023502.
XX 07-NOV-1995; 95US-0006315.
XX (CALJ ) CALGENE INC.
XX PA
XX PI Bugos RC, Rockholm DC, Yamamoto HY;
XX DR WPI; 1997-281036/25.
XX DR N-PSDB; AAT66243.
XX PT
XX DNA encoding plant violaxanthin de-epoxidase - used to modify the
XX sensitivity of a plant to light
XX PS Disclosure; Fig 3; 4lpp; English.
XX CC
XX The violaxanthin de-epoxidase (VDE) (AAW09876) of Arabidopsis
XX catalyses the de-epoxidation of violaxanthin to zeaxanthin and
XX antheraxanthin. This system, termed energy dependent non-radiative
XX energy dissipation or non-photochemical fluorescence quenching,
XX reduces the quantum efficiency of photosystem II (PSII), helping to
XX prevent PSII over-reduction and photoinhibitory damage. The amino
XX acid sequence of the VDE was deduced from an isolated cDNA clone
XX (AAT66243). VDE nucleic acids (see also AAT66241-42), in sense or
XX antisense orientation, can be used in genetic constructs to modify
XX VDE levels in plants. Increased levels result in the plant being
XX tolerant of increased light and therefore more productive and/or
XX more resistant to disease. Underexpression of VDE increases
XX photosynthetic efficiency under low light. The photosensitivity of
XX a range of crops, trees and ornamentals can be modified.
XX SQ Sequence 462 AA;

Query Match 32.2%; Score 2417; DB 18; Length 462;
Best Local Similarity 99.8%; Pred. No. 1.9e-180;
Matches 461; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 951 MAVATHCFTSPCHDRIRFFSDGIGRLGTRKRINGTFTLLKLPPIQSDADLTGGRSS 1010
Db 1 mavathcftspchdrirffsdgigrlgtrkringtftllkllppliqsadtltggrss 60

QY 1011 RPLSAFRSGFSGKIFDIVPLPSKNELKELTAPLLKLVLGVLACAFILVPSADAVDAKTC 1070
Db 61 rplsafrrsgfsgkifdivplpsknelkeltapllllklvglvacafilvpsadavdalktc 120

QY 1071 ACLKGCRIELAKCIANPACAAVACLOTNNRPDETECOIKGDDLFPNSVDFNECAV 1130
Db 121 aclkgrcielakciapacaanvacigtcnrrpdetecqikcgdlfensvdfnecav 180

QY 1131 SRKCVPRKSDLGFEPPADPSVLVQNFNISDFNGKWIITSGLNPTDFAPCQLHEFHEG 1190
Db 181 srkcvprksdlgfeppadpsvlvqnfnsdfngkwiitsglntpdaqcqlhefhteg 240

QY 1191 DNKLGVNLSWRIKTLDGFFTRSAVQKFPVQNPQGVLYNHDEYLYHQDDWYILSKIE 1250
Db 241 dnklgvnlswrktldsgfftrsavqkfpvqnpqgvlynhdneylhyqddwylsskie 300

QY 1251 NKPEYIFVYVRGNDWDGYYGAVVYTRSSVLVPSNIIPELEKAAKSIGRDFSTFIRTDN 1310
Db 301 nkpeyifvyvrgrndwdgygavvytrssvlvpsnlipelekaaksigrdfstfirtdn 360

QY 1311 TCGPEPALVERIEKTVEGERIIVKEVEEIEEVEEVEKEVEKVRTEMTLQRLAEGFNELK 1370
Db 361 tcgpepalveriektveegeriivkeveeieeVEEVEEVEKEVEKVRTEMTLQRLAEGFNELK 420

QY 1371 QDENFVRELKSEMEFLDETKMEASEVEKLFGLKALPIKVR 1412
Db 421 qdeenfvrelskeemefldekmeaseveklfgkalpikvr 462
```

RESULT 4
AAG4996
ID AAG4996 standard; Protein; 525 AA.
XX
AC AAG4996;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63310.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.

```
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148568.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.8%; Score 212.5; DB 21; Length 525;
Best Local Similarity 20.2%; Pred. NO. 4.5e-08;
Matches 139; Conservative 98; Mismatches 215; Indels 235; Gaps 36;

QY 270 FFTSAVQTFVODPDLPALYNHNEFLHYODDWILSSQIENKPDYIFVYGRNDAW 329
Db 1 ffamaaip--lkapslpqt-----recspflnrftvthqpvnrkfh----- 40
QY 330 DGYGGSVIYTRSPITLPIESIPNLQKAASVGRDFNNFITDNOSC---GPEPLVERLEKT 386
Db 41 -----psideasfs--irgrttvvr-----vletekstkiepepv--klial 80
QY 387 AEEGKLLIKEAVEIEEVEKEVEKVRDTEMTLFORLEGFKELQOQDENFVRELSKEEK 446
Db 81 vgkgevsplk-stsweevmlhtarrlkwwd-----egyemlvfdde---ivssndq 127
```



```
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153378.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.8%; Score 209; DB 21; Length 522;
Best Local Similarity 21.2%; Pred. No. 8.4e-08;
Matches 122; Conservative 76; Mismatches 191; Indels 186; Gaps 29;

QY 375 PEPPLVERLEKTAEGEKKLIKAEVETEVEEVEKEVEKVRDTEMTLQRLLEGKELQODE 434
D 68 pepvpv--klialvgkgevplk--stsevevmlhtarrlkvwd-----egvemlvfd 116
QY 435 ENFVRELSKEKETLNELOEATEVEKLFGRALPIRKLRNALAPHNSFLANHETIKYYVG 494
D 117 e-----llssndqtalt-lkqelbqtdllvvva-----vnnsesv----- 150
QY 495 SKLPGKRFSGWEDYFGSIVVAKISRRIPRYFRKSPRICGLDSRGLQLFSGKHNL 554
D 151 -----nwltqnsqvkumic-----fesspnlmarlvg----- 178
QY 555 SPAHSINQNPKNKGCKGKPKFDVALMV-----WEKMG-----QFAKTAIVAFILSV--- 601
D 179 tdvgsvnkd-----kevtevvktvedawetrnsddirfcollviinayirpvpvl 227
QY 602 ---ASKADAVALKTCCLLEKCEKLELAKICISNPACAAVACIQTCTNNRPDETECQKCG 658
D 228 qnlrsk-----gfstlscmvknegpqilncldpncrkalcinqcs--pvdqvcsvrci 280
```

```
QY 659 DLENSVDFNECAVSRKKCVPRKSDVGDGFPVPDPSPVLVQKPFDMKDFSGK----- 709
D 281 asyespyfeafslcvlqkncldakipvkpyvpv-----mtsfgrkelchdtaed 332
QY 710 -----WFIIRGLNPTFDADFDCOLHEHTEENK-----LVGNLSWR 744
D 333 lfvgwlgelewsrvvagnpaydqfpcyqlfyrfgkksfwyepvfvrtlegklvwr 392
QY 745 IRTPDGGFTTRSAVQKVFQDPKYP-----GILYNH--DNEYLLYODDWYILSSKVENSP 797
D 393 rr-----rysvkr-----gkipatfrfsvidngvvsnef-----wtivd--vsddls 432
QY 798 DYTFVYVYKGRNDWDGYSGLVLYTRSAVLP-ESIPELQTAQAQKVG-RDFTFIKTDTNC 855
D 433 wglfhyngaarvagqsygtavivtpdgsypaekekerlqsaalekcgikewelf-avdn 491
QY 856 GPEPPLVERLEKKEVERGER-----TIKVEEIEE 885
D 492 cenppl-----gipqgsrlhsrisieepseek 520

RESULT 7
AAG06330
ID AAG06330 standard; Protein; 526 AA.
XX
AC AAG06330;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 3066.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD 06-SEP-2000.
PF 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
```



```

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.78; Score 206; DB 21; Length 428;
Best Local Similarity 23.58; Pred. No. 1.1e-07;
Matches 85; Conservative 48; Mismatches 129; Indels 100; Gaps 18;

QY 582 WEKWC-----QFAKTAIVAIFILSV-----ASKADAVDAKLTCTCLKEGRLELAKCISN 631
DB 107 weRRnsddirfcIIvInayirpvlqnlrsk-----gfstlscmkvkcpgqilnclld 161

QY 632 PACAAVACLQTCNNRPDETCQIKGDLFNSVVDENECVAVSRKCVPRKSDVGDGPPV 691
DB 162 pncrkalkqlncqs--pvdqvcysrciasyespyfeafslcvlqkhncldeIdakipkpy 219

QY 692 PDPSVLVQKFDMDKDFSGK-----WFTIRGLNPTDADFCDQLHEF 730
DB 220 vpp-----mtsfrrgkelchdtaedlfvqwlgelewsrwrvaqnpaydqfpcqqlf 271

QY 731 HTEENK-----LVGNLSWRIRTPDGGFTFRSAVQKVPQPKYP-----GILY 772
DB 272 yrgkksfwyepvfqvtllegklvwr-----rysvkr-----gkipatfrfsvld 319

QY 773 NH--DNEYLLIQDDWYILSSKVENSPEDYIFVYKGRNDAWDGYGSGSVLITRSVLP-ES 829

```

```

Db 320 ngvvsnef-----wtlvd--vsddlswglfhyngaarvaqsytgavltvtpdgsypaek 371
QY 830 IPELQTAQRKVG-RDFTFTKTDNTCGPEPLVERLEKKVEEGER-----TIIKEVEEI 883
DB 372 ekerlqsalekcgikewelf-avdnscenppl-----gipqgsrlhstrisileepdse 424
QY 884 EE 885
DB 425 ex 426

RESULT 10
AAV44787
ID AAV44787 standard; Protein; 1055 AA.
XX
AC AAY44787;
XX
DT 04-MAY-2000 (first entry)
XX
DE Arabidopsis thaliana recombination repair protein, MIM.
XX
KW MIM; recombination; plant; DNA repair; hypersensitivity;
KW SMC protein family; Structural Maintenance of Chromosomes; MIM;
KW methyl methanesulphonate; irradiation; mitomycin C.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT Binding-site 49..56
FT Region /label= NTP_binding_domain
FT Region 184..442
FT Region /label= Coiled_coil_region-I
FT Region 443..627
FT Region /label= Hinge/spacer
FT Region 628..909
FT Domain /label= Coiled_coil_region-II
FT Domain 971..1007
FT FT /label= DA-box
FT FT /note= "conserved motif which harbours a Walker B type
FT FT NTP binding domain"
XX
WO200004174-A1.
XX
27-JAN-2000.
XX
14-JUL-1999; 99WO-EP04984.
XX
16-JUL-1998; 98GB-0015485.
XX
14-JAN-1999; 99GB-0000760.
XX
(NOV5 ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI Mengiste T, Paszkowski J;
XX
WPI; 2000-182437/16.
DR N-PSDB; AAZ50145, AAZ50146.
XX
New Arabidopsis polynucleotide encoding protein useful for assisting
PT recombinant repair of DNA damage in plants -
XX
PS Claim 1; Pages 22-25; 30pp; English.
XX
The present sequence is a MIM protein from Arabidopsis thaliana, which
CC contributes to recombination repair of DNA damage in plant cells.
CC The protein was tracked down with the help of a T-DNA tagged Arabidopsis
CC mutant showing hypersensitivity to methyl methanesulphonate (MMS). It
CC shows homology to a member of SMC (Structural Maintenance of Chromosomes)
CC protein family and confers hypersensitivity to treatment with MMS,
CC X-rays, UV light or mitomycin C. The present sequence is useful for DNA
CC repair in plant cells.
XX

```


QY 433 DEENFVRELS---KEEKEL-----449
Db 374 ereksiseldgkgeklilqrceetnayedsqykaageknslcleilnectsice 433
QY 450 ---NELOMEATEVEKLFGRALP--IRKRLMALAPHSNLANHEITIKYVVGSKLPCH---500
Db 434 nrknele---qlkeafakehgefiklafaearnqnlmleltvqqalrsemtdnqnns 489
QY 501 KRFSGWEDYFSGIVVAKICSSRRIPRYFRKSPRI-----CCGLDSRGLQLFSGHK 551
Db 490 kseagglkqelmtlkeeqnkmqkdvndllqeneqimkvmktkhecqnlesepir-----543
QY 552 HNLSPAHSINQVPGNSGCKPPKDVAlMVMKQGFQAKTAIVAlFILSVASKADAYDAL 611
Db 544 -----nsvkeresernq-cnfxpmdlevkeisldysnaqlvqleamrlnkelqlqese 596
QY 612 KTCTCL-----LKRELEL-AKGISNP-----ACAAN 637
Db 597 kekeclqhelqitrgdletsnlqdmqsqeisglkdceidaeeekyisgphelstsqndnah 656
QY 638 VAC-IQTCNNRPDETE--CQIKCGDLFNSVVDEFNECAVSRKKCV--PRK--SDVGF- 689
Db 657 lqcsigtmmklinelekeicellqaekye--lvtelnd---srsecitatrkmavevgkll 711
QY 690 ---PVPDPVSLVQKFDKMGKFWFTRGLNP-----TFDAFDC 725
Db 712 nevklindsglhbhelvedlp9gef---geqpneqhpvslapidesnsyehltisdev 768
QY 726 QLHFHEEENKLVNLSWRIRTPDGGFFTSVAVQKRVQDPKYPGILYNHDEYLLYOD-- 783
Db 769 qmhfaelqek-----flsqsehkildhqcq--mskmselqtvyds1 810
QY 784 ---DWILSKSVNSPEDYFVYVYKGRND-----AWDYGGSVLVTRSAVL 826
Db 811 kaenlvstlnrfngdvlkemqgleeglvpslssscvdpssslsgdssfy--rail 868
QY 827 PE---SIPELOTAQAQ---KVRDFNTFIKTDMTCGPEPL-----VERLEKKVEBGE 873
Db 869 eqtdmslslnlegavsanqcsvdevfssiqeenlrrketpsapakvveleslcevyr 928
QY 874 RTIIEVEIEEIEVEK---VRDKEVTLSKLFEG-----FKELQRDEENFLRELSKE 922
Db 929 qsl-----ekleekmesqgimknelelelellslerqelclrkqylseneqwqkltsv 984
QY 923 EMDVLDGLKMEATEVEKLF-----GRALPIRLKMAVATH-----CFTS-- 960
Db 985 tiemeskiaaekktqetqislevarlqlgldlssrsllgldtedaiggrnescdiske 1044
QY 961 -----PCHDRIRFFSSDDGIGRLGITRKRINGTEFLKILPPI-----QSADL--- 1002
Db 1045 htsettertphkd-vhqicdkdaqdlnidiekitetgalk---ptgecsqeqspdt nye 1100
QY 1003 -----RTTGGRS-----SRPLSAFRSGFSKGFIDIVPLPSK----- 1033
Db 1101 ppgedktgsgsciselsfsgpnaivpmdflngqedi hnlqlrvketsnenlrlhvwied 1160
QY 1034 -----NELKELTAPLLKLVGLVLAFLIVPSADAVDALTCACLLKGCRIELAKC 1084
Db 1161 rdrkvesllnemkeldsklhqevqlmt-----kieac-----ieleki 1199
QY 1085 IANPACAAVACLOTCNNRPDETECOIKCGDLFNSVVDEFNECAVSRKKCVPRKSDLG- 1143
Db 1200 v-----gel-----gel-----kk-----ensdlse 1212
QY 1144 ---EFPADPVSILVQNFNISDFNGKWYITSGLNPTFDADFQOLHEFHTE-----GDNKLGVN 1197
Db 1213 kleytsdhdqellqrvtse-----glns-----dlehmhadksredigdnvakvn 1258
QY 1198 ISWRIKTLDSGFFTSVAVQKRVQDPNQGVLYNHDNEYLH-----YQDDWILSKSTENK 1252
Db 1259 dswkerfid-----venelisrsekiasieahyleadlevvqtkclekclenhenk 1310

QY 1253 P-----EDYIFYYYRGNDWDGYYGAVVYTRSSVLPNSLIPELEKAAKSIGRDFSTFI 1306
Db 1311 qkvivcleelsvvtsernql---rgeldtmskttaldqlsekmkektqeshdsecl 1367
QY 1307 RTDNTCGPEPALVERIEIKTVEEGEIIIVKEVEEIEEE---VEKEVEKVGRTMTLFLQRLA 1363
Db 1368 hciqvaaevkektellqtlssdvselilkdtklqeklslekdskalsitkcelenqia 1427
QY 1364 EGRNELKQDENFVRE-----LSKEEMEFLEIK-MEASEVEK 1400
Db 1428 ----qlnkekellvkeseslqarlssesdyeklnvskaleaalvek 1468
RESULT 13
AAW23996
ID AAW23996 standard; Protein; 2482 AA.
XX AC AAW23996;
XX DT 28-MAY-1998 (first entry)
XX DE Human mitosis amino acid sequence.
XX KW Mitosis; phosphoprotein; mitotic cell cycle; antibody; analogue;
KW inhibition; M phase; Antagonist; hyperproliferative cell; cancer;
KW leukaemia; lymphoma; chromosome segregation.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT Domain 258..280
FT /note= "leucine heptad repeat"
FT Domain 340..362
FT Domain 564..593
FT Domain 1387..1443
FT Domain 1885..1962
FT Domain 2146..2188
FT Domain 2165..2187
FT /note= "leucine heptad repeat"
FT Misc-difference 2188
FT Misc-difference 2300
FT /label= "Bipartite targeting motif"
FT /note= "Optionally C or G"
FT Misc-difference 2189
FT Misc-difference 2301
FT Misc-difference 2303
FT /label= "Bipartite targeting motif"
FT /note= "Optionally A or T"
XX US5710022-A.
XX PD 20-JAN-1998.
XX PF 24-OCT-1994; 94US-0328254.
XX PR 24-OCT-1994; 94US-0328254.
XX PR 22-OCT-1993; 93US-0141239.
XX PA (TEXA) UNIV TEXAS SYSTEM.
XX PI Lee W, Zhu X;
XX WPI; 1998-109817/10.
XX N-PSDB; AAV09076.
XX New isolated mitosis protein and gene - useful for, e.g. developing
XX products for therapy and diagnosis of hyper-proliferative disorders
XX such as cancers or psoriasis
XX Claim 1; Column 40-52; 43pp; English.
XX This is the amino acid sequence for mitosis, a phosphoprotein
XX necessary for the cell to enter mitosis. The protein's degradation is
CC


```
Db 811 KAENLVLTNLRNFQGLVYKEMQGLGLEGLVPSLSSCVDPDSSSLSSGDSFY--RALL 868
* QY 827 PE-----SIPELQTAQ-----KVRDFTNTFKTDNCTGPEPPL-----VERLEKKVEEGE 873
Db 869 EOTGDMSSLNLEGAVSANCQSDEVCSSLOQENLTKRTPSAPAKGVVEELSLCEVYR 928
QY 874 RTIIVKEVEEIEEVEK---VRDEKVTFLSKLPEG-----FKELORDEENFLRELSKE 922
Db 929 QSL-----EKLEENESQGMKNKEIOLEQELSSERQELDCRKYVLSNEQWQKLTSV 984
QY 923 EMDVLDGLKMEAVEKLF-----GRALPRKLMVATH-----CFTS--- 960
Db 985 TLEWESKLAEEKQOTQOLSLEVARLQGLDLSRSSLLGDTDAIOGRNESCDSISKE 1044
QY 961 -----PCHDRIRFFSSDDIGRLGTRKINGTFLKILPPI-----QSADL--- 1002
Db 1045 HTSETTERPKHD-VHQICDKDAQDNLNDIEKITETGALK---PTGECSGEQSPDTNVE 1100
QY 1003 -----RTTGGRS-----SRPLSAFRSGFSKGIQDIVPLPSK----- 1033
Db 1101 PPGEDKTQGSSECSISELSESGPNALVPMDFLGNQEDIHNLQVRKETSNEENRLLHVIED 1160
QY 1034 -----NEKELTAPLLLKLVLGACAFILVPSADAVDAKTCACLLKGCRIELAKC 1084
Db 1161 RDRKVESILLNEMKELDSKLHLQEVQMT-----KTEAC-----IELEKI 1199
QY 1085 IANPACAANVACLOTNNRPDETCOIKCGDLFENSVDVDFNECAVSRKCKVPRKSDLG- 1143
Db 1200 V-----GEL-----KK-----ENSDLSL 1212
QY 1144 --EFPAPDPSVLQVFNISDFNGKWYITSGLNPTDAFDCOLHEFTE-----GDNKLGVN 1197
Db 1213 KLEYFSCDQHELLORVETSE-----GLNS-----DLEHADKSSREDIGDNVAKVN 1258
QY 1198 ISWRIKTLDSGFFTSVAVQKVFQDPNQPQGLVYNHDNEXLH-----YQDDWYILSSKIENK 1252
Db 1259 DSWKERFLD-----VENELSRIRSEKASTHEALYLEADLVVQTEKLCLEKDNEK 1310
QY 1253 P-----EDYIFVYGRNDWDYGGVAVYTRSSVLPNSIPIPELEKAKSIGRDFSTFI 1306
Db 1311 QKVIVCLEELS SVTSERNQL---RGELDTMSKKTALDQLSEKKMKETQOLESQSBL 1367
QY 1307 RTDTCGPEPALVERIEKTVGEERIIIVKEVEIEE---VEKEVEKVGRTMTLFORLA 1363
Db 1368 HCLQVAAEAEVKEKTELQTLSSDVSSELLKDKTHLQEKLOLESQALSIFKCELENOIA 1427
QY 1364 EGFENLQDEENFVRE-----LSKEEMEFIDEIK-WEASEVEK 1400
Db 1428 ----QLNKEKELLVKESESLOARLSSESDYEKLVNSKALEAALVEK 1468
```

RESULT 3

```
US-08-592-126-148
; Sequence 148, Application US/08592126
; Patent No. 5821091
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Transcripts Encoding Immunomodulatory
; TITLE OF INVENTION: Polypeptides
; NUMBER OF SEQUENCES: 151
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,126
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Rad50.pro-translation of SEQ ID NO:54
; US-08-592-126-148
```

Query Match

Best Local Similarity 1.8%; Score 135.5; DB 2; Length 1312;

Matches 124; Conservative 95; Mismatches 218; Indels 191; Gaps 24;

```
QY 382 RLEKTAEEGE----KLIIKEAVEIEEEVEKEVEKVRDTEMTLFLORLEGFKEQLQDEENF 437
Db 398 KLVREGEGETANQMLNDFAEKTELKQOIDEIRD-KKTGLGRIIELSEI----- 449
QY 438 VRELSEKEKEILNLOMEATEVEKLFGRALPIRKLRLMALPHSNFL-----ANHETIKY 491
Db 450 ---LSKKQ---NELKNVYELQLEGSSDRILELDQELIKAERELSKAENKSNVETLKM 502
QY 492 YVGSKLPGCHKRFSGWEDYFGSIIVVAKICSSRRIPRYFRKSPRICCCGDSRGLQFSGHK 551
Db 503 EVIS-LQNEK-----ADLRTLRK-----LDQEMEOL-NH-- 530
QY 552 HNLSPAHSINQNVPGNSGCKFPKDALVMYKMGWQFAPKTAIVAITLSVASKADAVDAL 611
Db 531 -----HTTQTQEMELTKDKADKDEQIRIKLSRHSDELTSLLGYF----PNKKQLEDWL 580
QY 612 KTCFCLLKECELELAKISNPACAANVACLOTNNRPDETECQIK-----CGDLFE 662
Db 581 HSKSKEINQTRDLAKL--NKELASSEQNKNHINNELKRREEQLSSYEDKLFQVCGSQDF 638
QY 663 NSVVDDEFNECAVSRKCKVPRKSDVGFVPDPSVLVQKQKDFSGKWFTRGLNPTFDA 722
Db 639 ESDDLRLKE-----EIEKSSKORAMLAGATAYVSQ 668
QY 723 FDCOLHEFHEEENKLVGNLSWRINTPDGFFTRSAVQKVFQDPKYPGILYNHDNEVLLYQ 782
Db 669 FITQL-----TDENQSCCPVCQV-----FQTEALQEVISD----- 700
QY 783 DDWVILSSKVENSPEDYIFV--YKGRNDWDYGGSVLYTRSAV-LPESIIPELQTAQ 839
Db 701 -----IQSKLRAPDKLKSTESSELKKEKRDEMGLVPMRQSIIDLKEIPELRNKLQ 755
QY 840 KVGKDFNTFIKTD-----NTCGPEPPLVERLEKKVGEERTIKE-- 879
Db 756 NVNRDIOR-LKNDIEEQETLLGTIMPEESAKVCLTDVTIMERFQMKELDKVERKIAQAA 814
QY 880 -----VEETEEVEKVRDKEVTLFSKLFEGFKELQDEENF-----LRELSKEE 923
Db 815 KLOGIDLDRTVQVQVNOEKQKHLDTVSSKIELNRKLIQDQEQEQIHLKSTTNELKSEK 874
QY 924 MDVLDGLKME-----ATEVEKLF 941
Db 875 LQISTNLQRQQLSEEQTVLSELSTEVSQSLY 902
```

RESULT 4

Query Match	1.8%;	Score 132;	DB 1;	Length 776;
Best Local Similarity	17.2%;	Pred. No. 0.017;		
Matches 139:	Conservative 117;	Mismatches 254;	Indels 3	

[illegible]

RESULTS

```

US-08-082-849B-2
: Sequence 2, Application US/08082849B
: Patent No. 5677274
: GENERAL INFORMATION:
: APPLICANT: Leppla, Stephen H.
: APPLICANT: Klimpel, Kurt R.
: APPLICANT: Arora, Naveen
: APPLICANT: Singh, Yagendra
: APPLICANT: Nichols, Peter J.
: TITLE OF INVENTION: Anthrax Toxi
: TITLE OF INVENTION: Related Meth
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townse
: STREET: Two Embarcadero Center

```

```
Db 511 AGYLENG-----KLQNRNIGLEIKDVQ----- 533
QY 1223 NQPGVLYNHDNEYLHYODDWYILSSKTIENKPEDYIFVYYGRNDWDGYYGAVVYTR--S 1280
Db 534 -----IIQSEKEYIRI--DAKVVPKSKIDTKIQE-----AQLNINQEWNKALGLPKYTKLIT 584
QY 1281 SVLPNSIPELEKAASIGRDFSTFIRDTNCGPEPALVERIEKTVBEGE--RIIVKEV-- 1337
Db 585 FNVHNRYSNIVESAYLILNWKNNIQSD-----LIKKVTNLYLDGNGRFVFTDITL 636
QY 1338 -----EIEEVEVEKEVKGRTMTLFLQRLAGFNLKQDENFVRE----- 1379
Db 637 PNIAEQYTHQDEIYEQVHSGLYPESRSILLHGPKSGV--ELRNDSEGFHEFGHAYDDY 695
QY 1380 ----LSKEEME-----FLDEIKMEASEV 1398
Db 696 AGYLLDKNQSDLVNTNSKKFIDIFKEGSNL 725

RESULT 7
PCT-US94-01624-2
; Sequence 2, Application PC/TUS9401624
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klompel, Kurt R.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; APPLICANT: Nichols, Peter J.
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
; STREET: Stewart Street Tower, 20th Floor, One Market
; STREET: Plaza
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01624
; FILING DATE: June 25, 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A.
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-115
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ. ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-01624-2

; Query Match 1.8%; Score 132; DB 5; Length 776;
; Best Local Similarity 17.2%; Pred. No. 0.017;
; Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

QY 733 ENKLIVGNLSWRIRTPDGGFTFRSAVQKFDVDPKYPGILYNHD--NEYLLYQDDWY--I 787
Db 72 EMYAIGG---KIYIVDGIYTHKHSLEASDSEKRRKIDYKGMALLHEHYVYAKEGPEY 128

QY 788 LSSKVENSPEDYIFVYYKGRNDWDGYYGGSVLYTRSAVLPSIIPPELOTAQKVGREDNT 847
Db 129 L---VIQSEDYVENTKALN-----VYETIGKILSRDILSKNQPYCKFLDLVLT 176
QY 848 FIKTONTG-----PEPPLVERLEKKVVEEGRTIIK----- 878
Db 177 IKNASDSGQDLFTNQKKEHPTDFSEVFELEONSNEVEVFAKAPAYYIEPQHRDVLQY 236
QY 879 -----EVEIEEVEKVRDKVTLFSKLPGEKE----- 907
Db 237 APEAFNYMDKFNQEINLSLELKDQRMLSRYEKWEKIKQHYQHWSDLSSEGRGLLKL 296
QY 908 ---LQDDENFURELSKEMOVLGDKMEATEV---EKLGRALPI-----RKL 950
Db 297 QIPIEPKDDIITHLSQBEKELLKRIQDSDFLSTEEKEFLKKLQIDIRLSSEEEKEL 356
QY 951 MAVATHCTFTSPCHDRIRFSSDDGIGRLGITRKIRNGTFLKILPPIQSD----LRTTG 1006
Db 357 L-----NRIQVDSSNP-----LSEK--EKEFLAKKLDIQPYDINQRLQDTG 396
QY 1007 GRSSRPL-----SAFRSGFSKIGFIDIVPLPSKNELKELTAPLKLVLG 1049
Db 397 GLIDSPSINLDVRKQYKRDQIQNIDALLHQSIGSTLYNKIYLYENNMINNLATL----- 450
QY 1050 VIACAFLIVPSADAVDAKLTCAKILKGRIELAKCIANPACAANVACLOTNNRPDETEC 1109
Db 451 -----GADLVDS----- 458
QY 1110 QIKGDLFENSVDDE--FNECAVSRKKCVPRK---SDLGEFPAPDPSPVLVONFNISDFNG 1164
Db 459 -----DNTKINGIFNEFKKFKYSSISSNYMIVDINERPALDNERLKWRIQLSPDTR 510
QY 1165 KWTITSGLMPTFDADFQCLHEPHEGDNKLV--GNISWRIKTLDSGFTFRSAVQKFDVDP 1222
Db 511 AGYLENG-----KLQNRNIGLEIKDVQ----- 533
QY 1223 NQPGVLYNHDNEYLHYODDWYILSSKIENKPEDYIFVYYGRNDWDGYYGAVVYTR--S 1280
Db 534 ----IIQSEKEYIRI--DAKVVPKSKIDTKIQE-----AQLNINQEWNKALGLPKYTKLIT 584
QY 1281 SVLPNSIPELEKAASIGRDFSTFIRDTNCGPEPALVERIEKTVBEGE--RIIVKEV-- 1337
Db 585 FNVHNRYSNIVESAYLILNWKNNIQSD-----LIKKVTNLYLDGNGRFVFTDITL 636
QY 1338 -----EIEEVEVEKEVKGRTMTLFLQRLAGFNLKQDENFVRE----- 1379
Db 637 PNIAEQYTHQDEIYEQVHSGLYPESRSILLHGPKSGV--ELRNDSEGFHEFGHAYDDY 695
QY 1380 ----LSKEEME-----FLDEIKMEASEV 1398
Db 696 AGYLLDKNQSDLVNTNSKKFIDIFKEGSNL 725

RESULT 8
US-08-404-531B-28
; Sequence 28, Application US/08404531B
; Patent No. 5863724
; GENERAL INFORMATION:
; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
; APPLICANT: Thomas, Gilbert Cole, and Robert Gagel
; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonyleurea Receptor
; Patent No. 5863724
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESS: No. 5863724ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/404,531B
; FILING DATE: 15-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: BYLR-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1498 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-404-531B-28

```

```

Query Match      1.7%; Score 124; DB 2; Length 1498;
Best Local Similarity 18.8%; Pred. No. 0.24;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

```

```

QY 868 KVEGERTIIKEVEIEEVEKVRDKEVTLFSLKLEGGFK--ELQDEENFLRELSKEEMD 925
   | : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 480 KLSQAQRTTL-----EYSNERLKQT---NEMLRGIKLLKYAWENIFCSVRKTRRK 528

QY 926 VLDGLKMEA--TEVEKLFGRLPIRKLMVATHCFTSPCHDRIRFFSSDD-----GIGRL 978
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 529 EMTSLRAFAVYTSISIFMTAIPAAVLITFV-----GHVSFFKESDFSVAFAFL 580

QY 979 GITRKRINGTFL-----KILPPIQ-----SADLRTTGGSSRPLSAFRSGSKG 1023
   | : | : | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 581 SLFHILVTLPLFLSSVVRSTVKALVSVQKLEFLSSAEIREEQCAPREPAQQA----G 636

QY 1024 IFDIVPL-----PSKNELKELTAPLLKLVGLACAFILVPSADAVDAKTCACLLKG 1076
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 637 KYQAVPLKVNKRKRPAREVRDLGLPLQ-----RLTPSTDG-DADNFCVQIIG 684

QY 1077 C-----RIELAKCIANPACAANVACLOTNNRPDETECOIKCGDLF 1117
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 685 FFTWTPDGIPTLSNITIRIPRGOLTMIVGVCGCKSLLLATLG-----EMQKVSQAVF 738

QY 1118 ENSVVDENECASVRKKCVPRKSDLGEPFAPDP-----SVLVQNFNI-SDF 1162
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 739 WNSLPD-----SEGRRPQOPRAGDSGRFCQEQPCGYASOKPWLNNATVEENITFESPF 793

QY 1163 NGKWI-----ITSGLNPTDFADQCQLHEPTEGDNKLVGNISWRIKT-----1204
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 794 NKQRYKMVIEACSLQPDIDL-----PHGDTQIGERGINLSTGGQRPDQCRPEPST 845

QY 1205 -----LDSGF-----FTRSAVKQFVQDPNPGVLYNHDNEXLVHYQDDWYILSS 1247
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 846 STPMIVFLDDPFSALDVHLSHLMQAGILELRDCKRTVVLVTKLQTLPHIA-DWIITAMK 904

QY 1248 KIENKPEYIFVYGRNDWDGAGVYVTRSSVLPNSIIELEKAASIGRDFSTFIR 1307
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 905 DGTIQREGTLKDFORSEQLFEHW-----KTLNMRQDQLEK-----ETVME 946

QY 1308 TDNTCGPEPALVERIEKVEGERIIVKEVEIEEVEKE 1347
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 947 RK--APEPS--QGLPRAMSSRDGLLDEDEEEEAASE 981

```

RESULT 9

```

US-08-476-900A-28
; Sequence 28, Application US/08476900A
; Patent No. 6031150
; GENERAL INFORMATION:

```

```

; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson
; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonyleurea Receptor
; Patent No. 6031150
; TITLE OF INVENTION: and Method of Detecting Persistent Hyperinsulinemic Hypoglycemia
; TITLE OF INVENTION: Infancy
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 6031150
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,900A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: BYLR-0027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1498 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-476-900A-28

```

```

Query Match      1.7%; Score 124; DB 3; Length 1498;
Best Local Similarity 18.8%; Pred. No. 0.24;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

```

```

QY 868 KVEGERTIIKEVEIEEVEKVRDKEVTLFSLKLEGGFK--ELQDEENFLRELSKEEMD 925
   | : : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 480 KLSQAQRTTL-----EYSNERLKQT---NEMLRGIKLLKYAWENIFCSVRKTRRK 528

QY 926 VLDGLKMEA--TEVEKLFGRLPIRKLMVATHCFTSPCHDRIRFFSSDD-----GIGRL 978
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 529 EMTSLRAFAVYTSISIFMTAIPAAVLITFV-----GHVSFFKESDFSVAFAFL 580

QY 979 GITRKRINGTFL-----KILPPIQ-----SADLRTTGGSSRPLSAFRSGSKG 1023
   | : | : | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 581 SLFHILVTLPLFLSSVVRSTVKALVSVQKLEFLSSAEIREEQCAPREPAQQA----G 636

QY 1024 IFDIVPL-----PSKNELKELTAPLLKLVGLACAFILVPSADAVDAKTCACLLKG 1076
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 637 KYQAVPLKVNKRKRPAREVRDLGLPLQ-----RLTPSTDG-DADNFCVQIIG 684

QY 1077 C-----RIELAKCIANPACAANVACLOTNNRPDETECOIKCGDLF 1117
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 685 FFTWTPDGIPTLSNITIRIPRGOLTMIVGVCGCKSLLLATLG-----EMQKVSQAVF 738

QY 1118 ENSVVDENECASVRKKCVPRKSDLGEPFAPDP-----SVLVQNFNI-SDF 1162
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 739 WNSLPD-----SEGRRPQOPRAGDSGRFCQEQPCGYASOKPWLNNATVEENITFESPF 793

QY 1163 NGKWI-----ITSGLNPTDFADQCQLHEPTEGDNKLVGNISWRIKT-----1204
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 794 NKQRYKMVIEACSLQPDIDL-----PHGDTQIGERGINLSTGGQRPDQCRPEPST 845

QY 1205 -----LDSGF-----FTRSAVKQFVQDPNPGVLYNHDNEXLVHYQDDWYILSS 1247
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Db 846 STPMIVLDDPFSALDVHLSHLMQAGILELLRDKRTVVLVTHKQLYLPHADWIIAMK 904
QY 1248 KIENKPEDYIFVYGRNDADWDGCGGAVVYTRSSVLPNSIPELEKAAKSIGRDFSTFIR 1307
Db 905 DGTQREGTLKDFORSECOLFEHW-----KTLNRDQOELEK-----ETVME 946
QY 1308 TDNTCGPEPALVERIEKVEGERIIVKEVEIEEVEEKE 1347
Db 947 RK---APEPS--QGLPRAMSSRDGLLDEDEEEEAASE 981
RESULT 10
US-08-488-546A-28
; Sequence 28, Application US/08488546A
; Patent No. 6054313
; GENERAL INFORMATION:
; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
; APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 6054313
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 6054313ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,546A
; FILING DATE: 07-JUNE-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/404,531
; FILING DATE: 15-MARCH-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: BYLR-0026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1498 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-488-546A-28
Query Match 1.7%; Score 124; DB 3; Length 1498;
Best Local Similarity 18.8%; Pred. No. 0.24;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;
QY 868 KVEGERTIIEVEIEEVEKVRDKEVTLFSLKFEFGK--ELORDENFURELSKEEMD 925
Db 480 KLSQAQRTTL-----EYSNERLKQT---NEMLRGTKLLKYAWENIFCSRVKTRRK 528
QY 926 VLDGLKMEA--TEVEKLFGRALPIKRLMAVATHCTSPCHDRIRFSSDD-----GIGRL 978
Db 529 EMTSRFAVNTSISIFNTAIPNAVLITV-----CHVSFFKESDESPSVAASL 580
QY 979 GITRKRINGTFL-----KILPPIQ-----SADLRTTGGRSRPLSAFRSFGSK 1023
Db 581 SLFHILVTPFLILSSVVRSTVKALYSVQKLSFEFLSSAIREEQCAPREPAPOGQA----G 636

QY 1024 IFDIYVPL-----PSKNELKELTAPLLLLKLVGLACAFILVPSADAVDALKTCACLLKG 1076
Db 637 KYQAVPLKVVNKRPAAREVRDLLGLPQ-----RLTPTSDG-DADNFCVQIIIG 684
QY 1077 C-----RIELAKCIANPACANVACLOTCNNRPDETECOIKCGDLF 1117
Db 685 FFTWPDGPTISNITIRPGOLTMIVGQVCGKSGLLLATLG-----EMOKVSGAVF 738
QY 1118 ENSVVDENECASRKCVPRKSDLGFEPPADP-----SVLVQNFMI-SDF 1162
Db 739 WNSLPD-----SEGRPOQPRAGDSRGFCOEORPCGYASQKWPWLLNATVEENITFESPF 793
QY 1163 NGKWY-----ITSGLNPTDADFQCLHEPHEFGDNKLVGNISWRIKT----- 1204
Db 794 NKQRYKMWIEACSLQPDIDL-----PHGDTQIGERGINSLTGGORPDQCRPEPST 845
QY 1205 -----LDSGF-----FTRSAVQKFQVDPNQPCVLYNHDNEYLHYODDNYILSS 1247
Db 846 STPMIVLDDPFSALDVHLSHLMQAGILELLRDKRTVVLVTHKQLYLPHADWIIAMK 904
QY 1248 KIENKPEDYIFVYGRNDADWDGCGGAVVYTRSSVLPNSIPELEKAAKSIGRDFSTFIR 1307
Db 905 DGTQREGTLKDFORSECOLFEHW-----KTLNRDQOELEK-----ETVME 946
QY 1308 TDNTCGPEPALVERIEKVEGERIIVKEVEIEEVEEKE 1347
Db 947 RK---APEPS--QGLPRAMSSRDGLLDEDEEEEAASE 981
RESULT 11
US-08-404-531B-6
; Sequence 6, Application US/08404531B
; Patent No. 5863724
; GENERAL INFORMATION:
; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
; APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 5863724
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 5863724ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/404,531B
; FILING DATE: 15-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: BYLR-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1581 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-404-531B-6

AND USES THEREFOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/357,014
FILING DATE: 19-Jul-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/574,959
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: DFN-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 905 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-357-014-9

Query Match 1.6%; Score 119; DB 4; Length 905;
Best Local Similarity 31.0%; Pred. No. 0.28;
Matches 39; Conservative 15; Mismatches 34; Indels 38; Gaps 4;
Qy 1284 PNSTIPELEKAAKSGRDFSTFIRTDNTCGPEPALVERI-----EKTVEEGERIIV 1334
Db 623 PPQLVPE-----CTPGGGP-PALEDLIVININSSDEEEEGE 665
Qy 1335 KEVEIEEEVEKEVKVGRTEMTLQRLAEGFNELKQDEENFVRELSKEEMFLDEIKME 1394
Db 666 EEEEEEEEEEEEE-----EEDFEEEDDEEYFEEEEEEEEEE 713
Qy 1395 ASEVEK 1400
Db 714 EGELEE 719

Search completed: November 6, 2001, 05:01:26
Job time: 7515 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 05:02:57 ; Search time 82.15 Seconds
(without alignments)
1309.294 Million cell updates/sec

Title: US-09-075-375A-4

Perfect score: 7495

Sequence: 1 MALSLHTVFLCKEALNLVA.....MEASEVKLFGLKALPIRKVR 1412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68: *
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2557	34.1	478	2 T03750	violaxanthin de-ep
2	2430	32.4	462	2 T00708	violaxanthin de-ep
3	210	2.8	522	2 A84606	hypothetical prote
4	194.5	2.6	2269	2 T28677	rhostry protein -
5	167.5	2.2	2401	2 T28676	hypothetical coile
6	158.5	2.1	1957	2 T38077	hypothetical coile
7	156	2.1	1365	2 T30822	Impl protein - Myc
8	155.5	2.1	885	2 H69378	conserved hypothet
9	155.5	2.1	2829	2 A42771	reticulocyte-bind
10	155	2.1	2469	2 H36812	hypothetical prote
11	151	2.0	1939	2 T18372	repeat organellar
12	148.5	2.0	1935	2 A59286	myosin heavy chain
13	146	1.9	1935	1 A37102	myosin beta heavy
14	146	1.9	2748	2 S57976	nuclear migration
15	144.5	1.9	839	2 S54174	DNA topoisomerase
16	144	1.9	2166	2 G70163	hypothetical prote
17	143.5	1.9	1558	2 B71603	RESA-H3 antigen pf
18	142.5	1.9	880	2 F75103	conserved hypothet
19	142.5	1.9	1199	2 T29145	hypothetical prote
20	142.5	1.9	1819	2 A71928	cas island protein
21	142.5	1.9	1935	1 S06006	myosin beta heavy
22	142.5	1.9	5105	2 T32650	hypothetical prote
23	141	1.9	1002	2 C70319	nitrite reductase
24	140.5	1.9	1302	1 JC6009	surface-located me
25	140	1.9	800	2 F64508	hypothetical prote
26	139.5	1.9	1025	2 S54044	probable membrane
27	139	1.9	2712	2 T05113	hypothetical prote
28	137.5	1.8	1875	2 S38173	myosin-like protei
29	137	1.8	1109	2 A40801	phosphoprotein pho

30	137	1.8	1676	2 E71410	probable centromer
31	136.5	1.8	3724	2 T18427	hypothetical prote
32	136	1.8	1127	2 T28317	ORF MSV156 hypothe
33	135	1.8	3660	1 S02041	dystrophin, muscle
34	134.5	1.8	1937	2 I38055	myosin heavy chain
35	134	1.8	978	2 A70387	conserved hypothet
36	134	1.8	1639	2 S05603	major merozoite su
37	134	1.8	2261	2 T20978	hypothetical prote
38	133.5	1.8	1934	2 I48153	myosin heavy chain
39	133	1.8	1624	2 T25592	hypothetical prote
40	133	1.8	1979	2 C71622	hypothetical prote
41	132.5	1.8	1156	2 B70356	chromosome assembl
42	132	1.8	809	1 JQ0032	anthrax toxin leth
43	131.5	1.8	1170	2 A72287	hypothetical prote
44	131	1.7	1156	2 E69444	chromosome segrega
45	130.5	1.7	1631	1 SAZQK1	major merozoite su

ALIGNMENTS

RESULT 1

T03750

violaxanthin de-epoxidase precursor - common tobacco

C;Species: Nicotiana tabacum (common tobacco)

C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000

C;Accession: T03750

R;Bugos, R.C.; Hieber, A.D.; Yamamoto, H.Y.

J. Biol. Chem. 273, 15321-15324, 1998

A;Title: Xanthophyll cycle enzymes are members of the lipocalin family, the first ide

A;Reference number: Z15054; MUID:98288256

A;Accession: T03750

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-478 <BUG>

A;Cross-references: EMBL:U34817; NID:gl1463122; PIDN:AAC50031.1; PID:gl1463123

A;Experimental source: strain Xanthi, tissue-type leaf

C;Genetics:

A;Gene: TWDEL

C;Function:

A;Description: violaxanthin de-epoxidase and zeaxanthin epoxidase catalyze the additi

d in protecting the photosynthetic apparatus from excessive light

A;Note: established as member of the lipocalin family

F;1-134/Domain: transit peptide (plastid) #status predicted <TNP>

F;135-478/Product: violaxanthin de-epoxidase #status predicted <MAT>

Query Match 34.1%; Score 2557; DB 2; Length 478;
Best Local Similarity 100.0%; Pred No. 4.4e-142; Mismatches 0; Indels 0; Gaps 0;
Matches 477; Conservative 0

QY 474 MALAPHSNFIANHETIKYVVGSKLPGRKFSWGWDYFGSIVVAKICSSRRIPRYFRKSP 533

Db 1 MALAPHSNFIANHETIKYVVGSKLPGRKFSWGWDYFGSIVVAKICSSRRIPRYFRKSP 60

QY 534 RICCGDLSRGLQFLSHGKHNLSPAHSTINQVPGNSGCKFPKDVALLMWEKWCQFAKTAI 593

Db 61 RICCGDLSRGLQFLSHGKHNLSPAHSTINQVPGNSGCKFPKDVALLMWEKWCQFAKTAI 120

QY 594 VAIFILSVASKADAVDALCTCTLLKECRLELAKCISNPACANVACLOTNNRPDETTC 653

Db 121 VAIFILSVASKADAVDALCTCTLLKECRLELAKCISNPACANVACLOTNNRPDETTC 180

QY 654 QIKCGDLFENSVDENECAVSRKKCVPRKSDVDFPVPDPSVLVQKFDKDFSGKWFIT 713

Db 181 QIKCGDLFENSVDENECAVSRKKCVPRKSDVDFPVPDPSVLVQKFDKDFSGKWFIT 240

QY 714 RGLNPTFDADCOLHEFHTEENKLVNLSWRITPDGGFTTSVAVKQFVQDPKYPGILYN 773

Db 241 RGLNPTFDADCOLHEFHTEENKLVNLSWRITPDGGFTTSVAVKQFVQDPKYPGILYN 300

QY 774 HDNEYLLYQDDWVILSSKVENSPEDYIFVYKGRNDWDGSGSVLYTRSAVLPSGIPE 833

Db 301 HDNEYLLYQDDWYLLSKVENSPEYIFVYKGRNDWDGCGSVLYTRSAVLPESTIPE 360
QY 834 LQTAAQVGRDNFTFIKTDNTGPEPLVERLEKKVEEGERTIIKEVEIEEVEKVRDK 893
Db 361 LQTAAQVGRDNFTFIKTDNTGPEPLVERLEKKVEEGERTIIKEVEIEEVEKVRDK 420
QY 894 EVTLFSKLFEGFKELQDEENFRELSEKEMVDLDGLKMEATEVEKLFGRALPIRKL 950
Db 421 EVTLFSKLFEGFKELQDEENFRELSEKEMVDLDGLKMEATEVEKLFGRALPIRKL 477

RESULT 2
T00708
violaxanthin de-epoxidase homolog F22013.3 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999
C:Accession: T00708
R:Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Con-
eologis, A.; Ecker, J.R.
submitted to the EMBL Data Library, April 1998
A:Description: Genomic sequence for Arabidopsis thaliana BAC F22013.
A:Reference number: Z14200
A:Accession: T00708
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-462 <SHI>
A:Cross-references: EMBL:AC003981; NID:g3063438; PID:g3063441; GSPDB:GN00059; ATSP:F22013
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: ATSP:F22013.3
A:Map position: 1
A:Introns: 72/3; 128/2; 160/3; 292/2

Query Match 32.48; Score 2430; DB 2; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.1e-134;
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 951 MAVATHCFTSPCHDIRFFSDDGIGRLGTRKINGTFFLLKILPPIQSADLRTTGRSS 1010
Db 1 MAVATHCFTSPCHDIRFFSDDGIGRLGTRKINGTFFLLKILPPIQSADLRTTGRSS 60
QY 1011 RPLSAFRSGFGKIPDIIVPLPSKNEKELKELTAPLLKLVGLVACAFILVPSADAVALKTC 1070
Db 61 RPLSAFRSGFGKIPDIIVPLPSKNEKELKELTAPLLKLVGLVACAFILVPSADAVALKTC 120
QY 1071 ACLLAGCRLELAKCIANPACAAVACLOTNNRPDETECCQKCGDLFENSVDDEFNECAV 1130
Db 121 ACLLAGCRLELAKCIANPACAAVACLOTNNRPDETECCQKCGDLFENSVDDEFNECAV 180
QY 1131 SRKCVPRKSDLGEPAPDPSPVLYQNFNISDFNGKWIYTSGLNFTTDAFCQLHEFHTEG 1190
Db 181 SRKCVPRKSDLGEPAPDPSPVLYQNFNISDFNGKWIYTSGLNFTTDAFCQLHEFHTEG 240
QY 1191 DNKLGNISWRITKLDGFFTRSAVQKFPQDPNPGVLYNHGNEYLHYQDDWYLLSKKIE 1250
Db 241 DNKLGNISWRITKLDGFFTRSAVQKFPQDPNPGVLYNHGNEYLHYQDDWYLLSKKIE 300
QY 1251 NKPEYIFVYVYGRNDWDGCGVAVYTRSSVLPNSIIPLEKAAKSIIGRDFSTFIITDN 1310
Db 301 NKPEYIFVYVYGRNDWDGCGVAVYTRSSVLPNSIIPLEKAAKSIIGRDFSTFIITDN 360
QY 1311 TCGPEPALVERIEKTEVEGERIIVKEVEIEEVEEVEKEVEKVRGRTMTLQRLAEGFNEK 1370
Db 361 TCGPEPALVERIEKTEVEGERIIVKEVEIEEVEEVEKEVEKVRGRTMTLQRLAEGFNEK 420

QY 1371 QDEENFVRELSEKEEFDEIKMEASEVEKLFGRALPIRKVR 1412
Db 421 QDEENFVRELSEKEEFDEIKMEASEVEKLFGRALPIRKVR 462

RESULT 3
A84606

hypothetical protein At2q21860 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84606
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: A84606
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-522 <STO>
A:Cross-references: GB:AE002093; NID:g4417279; PIDN:AAD20404.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2q21860
A:Map position: 2

Query Match 2.8%; Score 210; DB 2; Length 522;
Best Local Similarity 21.1%; Pred. No. 8.3e-05;
Matches 122; Conservative 77; Mismatches 186; Indels 194; Gaps 30;

QY 375 PEPPLVERLEKTAEEGKLLIKEAVEIEEVEEVEKEVEKVRDTEMTLQRLGFKELQODE 434
Db 68 PEPPLV-KLIALVKGGEVSPK-STSWEEVMLHTARLKWVD-----EGEYMLVFDD 116
QY 435 ENFVRELSEKEEILNLOMEATEVEKLFGRALPIRKLRMALAPHNSFLANHETIKYVVG 494
Db 117 E-----IVSSNDQRAMN-LFOELNQTDILVVA-----VNNSSEV----- 150
QY 495 SKLPGHKRESGWEDYFGSIVVAKICSSRRIPRYFRKSPRICCGLDRLGQLFSHGKHL 554
Db 151 -----NWIQTSKNVKNMIC-----FESSPLMNLGG----- 178
QY 555 SPAHSTNQNVKNSCKFPKDALMV-----WEKWG-----QFAKTAIVAFILSV--- 601
Db 179 TDVGSYNKD-----KEVEYVTVGDWERNRSDDIRFCLLLIINAIIRPVPL 227
QY 602 ---ASKADAVALKTCCLLKECRLELAKCIENPACAAVACLOTNNRPDETECCQKCG 658
Db 228 QNLRK-----GFSTLSOMVKNCGPOLLNCLLDPCRKALQCLNQS--PVDQVCSYRCI 280
QY 659 DLFENSVDDEFNECAVSRKKCVPRKSDGDFVPDPVSVLVQKFDKMDPSGK----- 709
Db 281 ASVEGPFYFEAFSLCVLQKHNCLLEDAKIPKYPVP-----MTSFRGKELCHDTAED 332
QY 710 -----WFTITRGLNPTFDADFQCLHEFH-----TEENKLVGN 740
Db 333 LFGWNLGELEWSRVVAGONPAYDQPCQYQLFYRGKKGKSSFWYEPVFOVRTLEKLV--- 390
QY 741 LSWRIRTPDGGFFTRSAVQKFPQDPKYP-----GILYNH--DNEYLLYQDDWYLLSSKVE 793
Db 391 --WERR-----RYSVKR---GKIPATREFSVLDNGVWSNEF-----WTIVD--VS 428
QY 794 NSPEDYIFVYKGRNDWDGCGSVLYTRSAVL-P-BSIPELOQTAAQKVG-RDFWTFIKT 851
Db 429 DDLWSGLFHYHGAARVAGOSYTGAVLVTPDGSYPAEKDKERLQSALEKCGIKEWELF-AV 487
QY 852 DNTCGPEPALVERLEKKEVEGER-----TIKEVEIEE 885
Db 488 DNGSCENPPL-----GIPQGRSLHSRISILIEEPDSEK 520

RESULT 4
T28677
rhopty protein - Plasmodium yoelii
C:Species: Plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28677; C4521
R:Keen, J.; Sinha, K.; Brown, K.; Holder, A.
Mol. Biochem. Parasitol. 65, 171-177, 1994

A:Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.

A:Reference number: Z20508; MUID:95021522

A:Accession: T28677

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2269 <KEE>

A:Cross-references: EMBL:L27838; MID:g457145; PID:g457146; PIDN:AAA21304.1

R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.

Mol. Biochem. Parasitol. 42, 241-246, 1990

A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple cd

A:Reference number: A45521; MUID:91101660

A:Accession: C45521

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 2131-2269 <KE2>

A:Cross-references: GB:M34283

Query Match 2.6%; Score 194.5; DB 2; Length 2269;
Best Local Similarity 18.4%; Pred. No. 0.0052;
Matches 255; Conservative 213; Mismatches 436; Indels 483; Gaps 65;

QY 304 YILSQIENKPDYIFVYVRGRNDWDGYSVYTRS-----PTLPESIPNLQ--KA 355

DB 34 YTISNQIKNLN--VSTYEGR---EGFTSSLELAKSWKTKLETITELTKSNEETVRL 87

QY 356 AKSVGRDNFTTDSNCGPEPLVERLEKTABEGL-----LIKEAVEI 401

DB 88 EKEIRELFKKY-----IDEEARKYLEGLKLELNKKIKDIIAKTEYKNTVEL 135

QY 402 EEEVEK-----EVEKVRDPEMTLFORLGFKELOQ--DEENFVRELSKEEK 446

DB 136 KKEIEKNAYIDELANQSPYKVTGYENKNTYIKSYFDQIYEGDIDTYNELSSIVK 195

QY 447 E-ILNELQMEATEVEKLGRLPI--RKLMLAPHSPFLANHET-----IKYVV 493

DB 196 EPDIDIE--DKTKLENLRSKIDNVYDKTQKMEIETVKSHLNINETNNKLPNTILEIKKYI 254

QY 494 GSKLPGHFRFGWEDYFGSIVVAKICSSRRIPRYFRKSPRICCGLDGRGLQFESH--- 549

DB 255 YDEI--SKELNKLMEFDKFNK---EKELSNKISDYDKRQEL--SEYKSKMLEIRHNYSQ 307

QY 550 -----GKHNLSPAHSINQNPV-----KNSGCGKFFPKDV 577

DB 308 TNVDNTKEEAKQNDKNEHMTTPTNEDELSKILSEVTKMDEILSKVNTYIDFNKY 367

QY 578 ALMWKKGQFAKTAIVAIFILSVASKADADVALKTCTCLLKECRLELAKCISNPACAA 637

DB 368 KETVNSHSQFTE-----LTDKIKAEVSDK-----ELKK----- 396

QY 638 VACLQTCNRP--DETECOIKCGDLFEN---SVVDFENECVSRKKCVPRKSD----- 685

DB 397 --CEOSFNDKSLINETNSIE--KEYONINTLKKVDEYIKVCKSTKESITKFSKOTIL 452

QY 686 -----VGDFFVPDPSVLVQFMKDFSGKWFITRGLNPNPFDAPDCOLHFFHTEE- 734

DB 453 KDLNQNLIKTVKETSIDKS-YIEKFE-QILTGK---QTKLENKFTESLNHANNEL 507

QY 735 -----NKLVGNIWSRIRTPDGGFFTSRVSQVKFVQDPKPYGILYHNHNEYLLODDMYILSS 790

DB 508 IKYFSDLANLGINENMLYNQFTEK--EKTFNDIKEKNI---HINEBI-----S 552

QY 791 KVNSPEYIF-----VYKGRNDWD-----GY 814

DB 553 KEIKIHASIYNSIEETEIREIGINIESLNTKVFVKENVTNLAKIKELKHDFSDFGK 612

QY 815 GSVLYTRS-----AV-----LPESIPELOATAQKVGDRDNFTFI 849

DB 613 EGNIKYTKIKKINDIMAVSQIDQHLNGLDDIQKSSYVSEMKQINKLEKVSNTFI 672

QY 850 KTDNTCG-----PEPPLVERLEK--VEEGERTIKEVEIEE---VEKVRDKEVTLFSLK 901

DB 673 SNDNVEGIKKKQIIVTKDKKKNIYEINKLLSELSIEKXDTSLKVKDINLSYQNL 732

QY 902 EEGFKELQRDENFLRELSKEEM---DVLGDKMEATEVEKLGRLPIRKIMAV--AT 955

DB 733 GNLFLE-QIDEEKKKAENTIKSMEAYIDDLNKKKSQEIETEMDIKMDINKEMALKIS 791

QY 956 HCTSPCHDIRFFSDDGIGRLGITRKRINGFTLLKILPPIQSADLRTTGGSSRPLSA 1015

DB 792 HDDDKKCHDKSK--NHKENISDIYKSSKIQDF-----SRESINDIKNKLQKNVSE 842

QY 1016 FRSGFS-----KGIFDIVPLPSKNEKEL-----TAP 1042

DB 843 SQNHSDINOCLENAVINYILKL---NKIKIIDKVKEYTSEIEKNKNINDELANSSEK 899

QY 1043 LLLKLVGLACAFILVPSADVALKTACCLGKRIELAKCIAN---PACANVACLOT 1099

DB 900 VIKKIEGDLN-----LAECKSKINSTLDDKDIDECININVLAK 938

QY 1100 CNRNPDETEC--QIKCGDLFENSVDDEFNRC----- 1128

DB 939 -NILNEETNITNHFKAEEYKIVLSNFNMIEMADNKSQVILEIKKNGTNDHDYNIKEL 997

QY 1129 -----AVSRKKCVPRKSDLGEPAPDPSPVLVONFNISDFNGKWTSGINPT 1175

DB 998 KSHKDSNGYKTEADQNKKAIOKNKELFEQYKEEVTLLNKYYAVELKKN- 1047

QY 1176 FDADFQCLHEFHTEGDNKLVGNISWRIK-----TLDSGFFTSRVSQVKFVQDPNQPGVLY 1229

DB 1048 ---FD-----KYNKSK---QIIEKIDAHNYCTLESG-----KSEKKNE- 1082

QY 1230 NHDNEYLHYODDYILSSKIENKPEYIFVYVRGRNDWDGYSVYTRSSVLP----- 1284

DB 1083 -IKNEKIHIDE--VANNDKSNK-----AITSIKVSVEPFKTI 1118

QY 1285 ---NSI-----IPELEKAAKSIGRDFSTFIRDTNCTGPEPALVERIEKTVVEGER 1331

DB 1119 IKINEIRTKSDCLKETNDLEKQISNLSIDTQETKLTEN--GKOLKTLLELLESLAKKQ- 1175

QY 1332 IIVKEVEIEEVEKEVEKVGRTMTLQF-----LAEGFNKLKODEENFY---RELS 1381

DB 1176 ---KNIEDQKKELDENVSKIKNIENVTNOHKKYEIGIVEKINEIAKTNKQIESTKELI 1232

QY 1382 KEEMEFL 1388

DB 1233 KPTIQHI 1239

RESULT 5

T28676

rhoptry protein - Plasmodium yoelii (fragment)

C:Species: Plasmodium yoelii

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000

C:Accession: T28676; A45521

R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.

Mol. Biochem. Parasitol. 76, 329-332, 1996

A:Title: Comparison of two members of a multigene family coding for high-molecular ma

A:Reference number: Z20507; MUID:97077455

A:Accession: T28676

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2401 <SIN>

A:Cross-references: EMBL:U36927; MID:g1041784; PID:g1041785; PIDN:AAB41263.1

R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.

Mol. Biochem. Parasitol. 42, 241-246, 1990

A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple

A:Reference number: A45521; MUID:91101660

A:Accession: A45521

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 2260-2401 <KEE>

A:Cross-references: GB:M34281

Query Match 2.2%; Score 167.5; DB 2; Length 2401;

[illegible]

Db 457 IKDFE---KTEODRACLSSSSNELKESALIDKKQDELNNLREQIKQKKVSESTQSSL 513
QY 277 QTVQDPDLPALYNHNDNEFLHYQDDWYLLSQIENKPPDYIFVYVYGRNDADWDYGGSV 336
Db 514 QSLQD-----ILNEKKKHEVYESQLNELKGLQTE----- 544
QY 337 IYTRSPTLPSIIPNQAASVGRDFNNFITDSCGPEPLVERLEKTAEGEKKLIK 396
Db 545 -----ISNSEHSSQSLLAAEKEAAVATNE-----LSEKNSLOT 581
QY 397 EAVEIEEVEKEVKVRODEM-----TLFORLLEFGKELQOEDENFVRELSEKEKEILN 450
Db 582 LCNAFOELAKSVMLKENEQNFSSLDTSFKKLNEHQELENHQTITKQL-KDTSSKLQ 640
QY 451 ELQMEATEVEK-----LFGRALPIRLMALAPH-----SNFLANHETIKYVGS 495
Db 641 QLOLERANFEQKESTLSDNNDLRTKLKLEESNLSLKQEDVDLSLEKNTQTLK----- 695
QY 496 KLPGRKRFSGWEDYFGSVIVAKICSSRIIPRYFRKSPRICGLDSRGLQLF---SHGKH 552
Db 696 -----ED-----LRKSEALRFSK-----LEAKNLREVIDNLKXKH 726
QY 553 NLSPA-----HSI-----NONVPKNSGCKPKPDVALMVKWQGFATAIIVAFILSVA 602
Db 727 ETLEAQRNDLHSSLSDAKNTAILSELTKSSEDV-----KRLTANVETLTQD 774
QY 603 SKA-----DAVDALKTCCLLKECELELAKICISNAPCAANVACLOTNNRPDETECOIK 656
Db 775 SRAMQSFSTSLVNSYOSISNLYHELRDHHN-----MOSQNTLLSEESKLK 821
QY 657 --CGDLFEN--SVVDEF-----NECAVSRKKCVPRK-----SDVGDPPVDP 694
Db 822 TDCENLTQONMTLIDNVOKLHMKHYNOESKYSSELKVLGLSLDLKLNRLSLNVAISND 881
QY 695 SVLVQKFDK-KDFSGKWFITRGLNPTTFDAFDCQLHEFTTEE-----NKLVGNI-SWRIR 746
Db 882 QILTQALAEISKNYDLEQESAQLNSGLKSLAEKQLLHTENBELHRLDKLTGKLGK---- 937
QY 747 TPDGFFFTSAVQKVFQDPKPYGILYNHNDNEVLLYQDDWYLLSSKVENSPEDYIFVYK 806
Db 938 -----IESKSSDL-----GKLTARQEE--ISLNKEENMSQAITSVKS 976
QY 807 RNDADWDYGGSVLYTRSAVLPSIIPELQTAQKGRDFNTFIKTDNCTCGPEPLVERLE 866
Db 977 KLDE-----TLKSSKL-EADIEHLKNKYSEVEVERNALASN-----ERLM 1017
QY 867 KQVE-EGERTIIKEVEEIEEVEKEVKRDEVILFSLK-----FEGF-----KELQD 911
Db 1018 DDLKNGE-----NIASLQTEIEKKRAENDDLQSKLSVVSSEYENLLLISSQTNKSL-D 1071
QY 912 EENFLRELSKEBMDVLDGLKMEATEVEKL--FGR-----ALPIRLKMAVATHC 957
Db 1072 KTNQKYITEKNVOKLLDEKQDORNVLEBELTSKYKLGGENAQIKDELLALRKK----- 1124
QY 958 FTSPCHDRIRFSSD-----DGIGRL-----GITRKRINGTFLILKPIIQS--AD 1001
Db 1125 -SKQOHDLCANFVDLKEKSDALEQLTNEKNELIVLSLQSSNNEALVEERSDLANRLSD 1183
QY 1002 LRTTGGRRSRPLSAFSGFGIFDIVLPKSNKELKELTAPLLLLKLVGLACAFIIVPSA 1061
Db 1184 MKKLSDSQDNVIVIRSD-----LVRVN 1206
QY 1062 DAVIDALKTCACLLKGRTELAKCIANPACAAANVACLOTNNRPDETECOIKCGDLFENS 1121
Db 1207 DELDTLK-----KXDSLSISTQYSEVCQRDRLDLSLKGK----- 1240
QY 1122 VDEFNECAVS-RKKCVPRKSDLGEPPAPDPSVLQVN--FNISDFNGKWIIT-SGLNPTFD 1177
Db 1241 EESFNKYAVSLRELCTKSEIDV-----PVSEILDNDNFVNAGNFSLSRLVLSLENVLD 1295
QY 1178 AFDCQLHEFTGDKNLVGNISWRITKTLDSGFFTRSAVQKVFQDPNQGVLYNHNDNEYLH 1237
Db 1296 AFN-QVNFKKMELDN-----RLTTTDAEF-----TKVADL-----EKLOH 1330

QY 1238 YQDDWYLLSSKIENKPEDYIFVYVYGRNDADWDYGGAVVYTRSSVLPNSIPIPELEKAAS 1297
Db 1331 EHDDWLI-----QRG-----DLEKALKD 1348
QY 1298 IGRDFTSFIRTDNCTGPEPALVERIKTEVEGERIIIVKEVEEIEEVE----- 1345
Db 1349 SEKNF-----LRKEAEMTENIH-SLEGKEETKKEIAELSSRLDQNLATNKLKNO 1398
QY 1346 -----KEY-----EKVGRTEMTLFORLAEGFNLKQDEENFVRELSEKEMEFL--DEIKMEA 1395
Db 1399 LDHLNQELRLKEDVLKESLIISLEESLSNQKQESSLLD--AKNELEHMLDTSRKNS 1456
QY 1396 SEVEKL 1401
Db 1457 SLMEKI 1462

RESULT 7

T30822

Impl protein - Mycoplasma hominis

C/Species: Mycoplasma hominis

C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 07-Dec-1999

C/Accession: T30822

R/Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christiansen, G.

Infect. Immun. 63, 3336-3347, 1995

A/Title: Selection of Mycoplasma hominis PG21 deletion mutants by cultivation in the

A/Reference number: Z18884; MUID:95369882

A/Accession: T30822

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: DNA

A/Residues: 1-1365 <JEN>

A/Cross-references: EMBL:U01962; NID:g790243; PID:g790244; PIDN:AAA81013.1

C/Genetics:

A/Gene: Impl

A/Genetic code: SGC3

Query Match 2.1%; Score 156; DB.2; Length 1365;
Best Local Similarity 16.8%; Pred. No. 0.46;
Matches 209; Conservative 187; Mismatches 420; Indels 430; Gaps 52;

QY 272 TRSAVQTFVQDPDLPALYNHNDNEFLHYQDDWYLLSQIENKPPDYIFVYVYGRNDADWD 331
Db 282 TRNQIQEFINT-----NKNP--NYSE---LISQLTSKRD-----SKNSVYDS 319
QY 332 YGGSVIVYTRSPTLPSII-PNLOKA-----AKSVGRDFNNFITDSCGPEPLVERLEK 385
Db 320 SNKSDIESANTELKQALAKANADKQVADNLAKSIKEIQLNNSVSNANT-----LSA 369
QY 386 TAEGEKLLIKEAVEIEEVEKEVEKYRDTMTLPQLLEGF-----KELOQDEENFVREL 441
Db 370 KLTDKNTIOQAATELEKEVQKQDAIKSNNTASMQSAKSLDAKVAEITTKLETFF----- 425
QY 442 SKBEKELTNELOMEATEVEKLFGRALPIRLMALAPHSNLANHETIKYVYVSKLPGHK 501
Db 426 NKQKEAFNELKQTRNQIQEF----- 446
QY 502 RFSWGVEDYFGSVIVAKICSSRIIPRYFRKSPRICGLDSRGLQLFSPHGHKNSIPAHS-I 560
Db 447 -----INTNKNPNYSELISQLTSKRSKNSVTDSSNKSDESANTEL 489
QY 561 NONVPKNSGCKPKPDVALMVKWQGFATAIIVAFILSVASKADADVADKTCCTLLKE 620
Db 490 QKALAKAN-----ADKVQA-----DRLAKS 509
QY 621 CRLELAKICISNAPCAANVACLOTNNRPDETECOIKCGDLFENSVDVECEAVSKKCV 680
Db 510 IKEQLNNSVSN-----ANTLSAKLTQDN--TIQAKTE 541
QY 681 PRKSDVGDFFVPDPSVLVQKDFMKDFSGKWFITRGLNPTFD-----FDCQLHEFTHEENK 736
Db 542 LEKE-----VOKADQAIKSNNTASMQSAKSLDAKVAEITTKLETFFNKDKEA 588

C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

```

QY 737 LVGNLSWRIPTDGGFFTRSAVQFVQDPKYPGILYNHNDYLLYQDDWYILSSKVENSP 796
Db 589 KFNELKO-----TRNQIEFINTNK-----NNPN-----YSELISQUTSKRDS-- 626
QY 797 EDYIFVYVYKGRNDWDGCGSVLYTRSAVLPESIIIP-----ELQTAAGVGRDFNTFIK 850
Db 627 -----KNSVTSSNKSDDIESANTELKQALANADKQVADNLAISIREQLNNSVS 676
QY 851 TDNCGPEPLPVERLEKKVEEGERTIIEVEEIEEVEK-----VRDKVETLF 898
Db 677 NANT-----LSAKLTDKNTIQQAQTELEKEIKQANOAISNNTASMOAKSSILD 726
QY 899 SKLPEGFKELO---ROENFLRELSREEMDVLDGLKMEATEVEKLFGRALPIRKLMVAT 955
Db 727 AKVAEITTKLETKNKEAFNELKOTRNOI-----QEFINTNK----- 765
QY 956 HCTSPCHDRIRFFSDGIGRLGTRKINGTFLKILPPIQSA---LRTTGGRS 1009
Db 766 ---NNPNYSEL-----ISQTSKRSKSNVYTDSSNKSDDIESANTELKQALNTAKAKK 814
QY 1010 S-----RPLSAFSGSKGIDIVLP-----PSKNEL-KELTAPLLKLVG 1049
Db 815 SSIDNELRPL---KNDLQSKIEEFGPIRNTNFSWISSKLETTKNKLAELT----- 862
QY 1050 VLACAFILVPSADAV-----DALKTCACLLKGRICELAKCIANPACAANVACLOTGN 1101
Db 863 -----KADAINKNPSSQKALNDSSQVQKGLNELLKITE-----EPGKVETKN 907
QY 1102 NRDPETEIQKCGDLFNSVDFNECAVSKKCVPRKSDLGFEFPAD-----PSVLVQ 1155
Db 908 SNIGYRLFLKAQAFQFNNSVDVLLKNAWEKQTLSSKQKLGNOSTKDYLTQLSTEMSTQ 967
QY 1156 NFNISD--FNGKWIYITSLNPTDAFCQHEPHTEGDNKLVGNISWRITKLDGSGFFTR- 1212
Db 968 ESTTKKVVIVIAQHIRNLN-----SOYRLEAD-KLIAN-----MKRGYGDV 1009
QY 1213 --SAVQFVQDPNPGVLYNHDNYLHYQDDW-----YILSSKIENKPEYIFVYVGRN 1265
Db 1010 GIESLQKW-QDLMDSDVLSVDDS-----LKDDFNKALRVLGVDYTKPNPVSSWFKNRNS 1064
QY 1266 -DAWDGCGAVVYTRSSVLNSI-----IPELEKAASIGR----- 1300
Db 1065 IENTQNLRLILVRENIILDKARDLKRAEKTIFVDENINSLDQRAKRLKEILNKN 1124
QY 1301 DFTSFIR-----TDNCGPEALVERIEKTVVEGERIIVKE-----VBEIEEVEKEV 1348
Db 1125 DLSNFTLNHQNKQFTAKDITPKISLLENKLINEINQYLLPIKEKAVSKISEIEKN-KKEL 1183
QY 1349 EKVGRTEMTLQRLAEGFNLKODEENFVRELSEKMEFDEIKMEASEVEKLFQKA 1405
Db 1184 EDIIRSNFYLWE-----KVEINKYIISLTNKKQVELRNSINFE 1220

RESULT 8
H69378
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Feb-2001
C:Accession: H69378
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiaach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343
A:Accession: H69378
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-886 <KLE>
A:Cross-references: GB:AE001032; GB:AE000782; NID:g2689355; PID:AAB90211.1; PID:g264956

```

Query Match 2.18; Score 155.5; DB 2; Length 886;
Best Local Similarity 20.6%; Pred. No. 0.27; Indels 221; Gaps 27;
Matches 137; Conservative 85; Mismatches 221; Indels 221; Gaps 27;

```

QY 797 EDYIFVYVYKGRNDWDGCGSV--LYTRSAVLPESIIIPQLQTAAGVGRDFNTFIKTD-N 853
Db 152 EDY-----ENAWKNLGAVIRMLEREKERKLEFLSQEEQIKQKEK-----KAEIE 197
QY 854 TCGPEPLPVERLEKKVEGERTI---IKEVEIEEVEKVKDEVTFLSKLFEGFKELQR 910
Db 198 RISEIKSIESIRLEKLSSEVRNLESRLKELEHKSRLKQESSVLQEV-----R 249
QY 911 DEENFLRELSREEMDV---LDGLKMEATEVEKLFGRALPIRKLMVATHCTSPCHDRIR 967
Db 250 GLEEKLRLEKQLEKVEVRIEDLEKKAKEKEL-----KPKAERY- 289
QY 968 FSSDDGIGRLGTRKINGTFLKILPPIQSA---LRTTGGRSRPLSAFSGSKGIF 1025
Db 290 -----SILEKLLSEINQALRDVEKREGDLTREAAAGTQAOLKAAEE 329
QY 1026 DIVPLPSKNELKELTAPL-----LKLVLGVLACAFILVPSADAVDALKTC 1070
Db 330 D-----NSKLEETIKRTEELERELERPEKSHRLLE-----TLKPKMDRMQGIK-- 372
QY 1071 ACLLKGRIELAKCIANPACAANVACIQTCCNRDPETEIQKCGDLFNSVDFNECAV 1130
Db 373 -----AKLEEKNLTPDKVE---KMYDILLSKAEKEE-KEITE 404
QY 1131 SRKKVPRKSDLGFEFPAPPSVLVQNFINISDFNGKWIYITSLNPTDAFCQHEPHTEG 1190
Db 405 KKLKLIJAKSSLTGAGLKKAAVEE-----LKSAERTCPVCGRELDDEHRKN 451
QY 1191 -----DNKLGNISWRITKLDGSGFFTRSAVQFVQDPNPGVLYN-- 1230
Db 452 IMAEYTRMKRTAEELAKADEIEKKERLEKVEKALEKETVLKYRQMYDELKALANEL 511
QY 1231 --HDNEYLHYQDDWYILSSKIENKPEYIFVYVGRNDWDGCGAVVYTRSSVLNPSII 1288
Db 512 SSHDAE-----KLSAEE-----YRKVERLDGLRQOKILLSSA---SRI 550
QY 1289 PELEKAASIGRDFSTFTDNTDNTGPEPALVERIEKTVVEGERIIVKEVEIEEVEKEV 1348
Db 551 KELKSLREI-----EEAL-KNVESERGELHRIKIREEGFESLELEPREV 593
QY 1349 EKVGRTEMTLQRLAEGFN---ELKODEENFVRELSEKMEFDEIKMEASEVEKLFQKA 1405
Db 594 -----QSLRPFTYKWLKDAESRLSESELKRRE-KLEDEISEIAKLEANGKA 641
QY 1406 LPIR 1409
Db 642 EIR 645

```

RESULT 9
A42771
C:Species: Archaeoglobus fulgidus
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 28-Apr-1995
C:Accession: A42771
R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992
A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
A:Reference number: A42771; MUID:92315338
A:Accession: A42771
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2829 <GAL>
A:Experimental source: Belem strain, merozoites
A>Note: sequence extracted from NCBI backbone (NCBIN:108114, NCBIIP:108115)

Query Match 2.1%; Score 155.5; DB 2; Length 2829;
Best Local Similarity 17.4%; Pred. No. 1.4;
Matches 283; Conservative 242; Mismatches 573; Indels 529;

Qy	57	TSYKTSFSFSDSHCKDKSOICSDIDSFP	-----QRFDLKRQMTLLEKQWRQFTOLA	110
Db	1110	TSNKNEILKSVKEVEDKLNVLQFQEDYK	KKVKNPENKEQLEAIRGSMKLAB-----	1160
Qy	111	IVLVCFTVIVPRDVA	DALKTCACILKECRTELAKCIANPSCAAVACLOTNNRPDETE	170
Db	1161	-----VINKHVSEMTQLESTANTLK	-----SNAKGKNEHDLBELN-----KTK	1119
Qy	171	CQIKCGDLFE--NSVVQDFNECAVGRKK	CVPRKSDVGGFPYDPDRNA-----VVQNFNM-KDF	224
Db	1200	GQMR--DIYEKLLKIAEELKEGTVNELD	DANEKANKEV-PEPERNIIGHVLERITVEKDK	1256
Qy	225	SGWYITSGINPTFDADFQOLHEF--H	MENDKLVGNLTWRIKTLDGGFFTRSAVQTVQD	282
Db	1257	AGK--VVEEMSNLTKIEKLQIETSD	SDSNLVTTSTIKHLENAK-----	1300
Qy	283	PDLGALYNHNEFLHYQDDWYILSSQ	TENKPDYFVYVGRN-DAWDGVGGSVITRS	341
Db	1301	-----YED-	-----VIKRNEESIQUREKASLETLD-----EM	1328
Qy	342	PTLPESIIPNLQAAK--SVGRDFNN	-----FITDINS-----	372
Db	1329	KLVOQVMNLQSAIQGNAGISKELNEL	KGVIELLISTNYSSILEYVKKNSSSVRFSQL	1388
Qy	373	CGPEPPLVERLEKTAE--EGEKLLI	KEAV-----EIEEVEK-----EVEKVRDT	415
Db	1389	ANGEFTKAEGEKNASARLAEAEKLE	--REQIVKDLDYSDIDDKVKKEGTREILKMKES	1446
Qy	416	EMTLFORLEGFKELQOODENFR	ELSKKEILN-----ELQME-----	455
Db	1447	ALTFWEE-SEKFKQMCSSHWEKAK	EGKKIEYLVKNNGDGGKANITDSOMEVGVNYSKAE	1505
Qy	456	-----ATEVEKLFGRALPIRKL	RLMALAPHNSFLIANHETIKYY	492
Db	1506	HAFHTVEAQDKTKAFCE	SIAYVTKMNLFNESL-MKEVVKCEK-----KND	1558
Qy	493	VGSKLPGHRRFSGWEDYPGS	IVAAKICSSRRIPRYPRKSPRICGLD	1603
Db	1559	-SAKL-----KPYDGR--KARV	SENERKISLKEAKV--PKKSSQUNDVSTK	1603
Qy	553	NLSPAHSINQNPVKGNGCK	FKFDVALMWKWKQGFATAIVAIFILSVASKADAVDALK	612
Db	1604	SLAQIDNCRQOLDVLSNIGRV	KQNALQYFDS---ADKSMKSVLPISLGA	1659
Qy	613	TCCLLKECRELELAKCISN	PACAANV--ACLOTNNRPDETECO-ITCGDL	1669
Db	1660	AA---KESYENLETVMQEM	SRIINVEEGSLTDIDDKTTIENDLLKMKOYE	1715
Qy	670	NECAVRKKCVPR-KSDVGD	FPVPDPSVLQKDFGKWFITRGLNPTFDADFQOLH	728
Db	1716	KENADRKSNFELVGSEIN	ALLDPSTSI-FI-KLKLKEYD-----WTGDK	1766
Qy	729	EPFTEENKLVGNLSMRI	-----RTPDGGFFTRSAVQK	760
Db	1767	EIRGEFTKSYNLIETHLS	NATDYSVTFEKAQSLRELAKEEHLRRREER	1826
Qy	761	FVQ-----DPKYPGILYN	DNEVILYQDDWYILS-----	789
Db	1827	VESLKLLKEMKKVSAEY	EGMKRDHTSVSQLVDQDKMTITVDELKTLNDISE	1886
Qy	790	--SKVENSPE-DYIFVY	YGRNDAMDGVGGSVLYITRSVLPESIIPELQ	1846
Db	1887	IVKKVRESKHADY-----	RRDANSWYESMYTLANYFLSDEAKISS-----	1931
Qy	847	TFTKTD-----NTCGP	PEPLVERLE-----KKVEEGR-----TIKEY	880
Db	1932	AEKSNFKTDLLEIFSVIS	NSNELLKTIQSDNDVIQKRESOLAKDADIVNVIK	1991

Qy	381	ETEETEEVKEVYRDEKVFILPSKLPFGPKELORDE-----ENFLRELISKEEMDVLDGLKWEA	934
Db	1992	NEFEKLEEAKEEVVSEKVEALRKSQVEGIRCFHFENFHLNDTNE-----	2040
Qy	935	TEVEKLFGRALPRLKLMVAVATHCFSPCHDRIRFFSSDDGIGRLGITRKIRINGTFLFKIL	994
Db	2041	ELENL-----KKVVTIY-----RDKSERESGLQEM-----ENENN-TYSNSI-	2077
Qy	995	PPTQSADLRTTGGRRSRPLSAFSGFSKGIPIVPLPSKNE-----LKELTAPILLKLKLVG	1050
Db	2078	--TQLEGIVVSAGESKE-----DIEKLEERSNEEMRNISEKISTIDSK----	2117
Qy	1051	LACAFLLVPSADAVDAKTCACLLKGCRTELAKACTANPACAANVACLOTGNRRPDETEQ	1110
Db	2118	-----VIENNSTIDELK---LGRNQAHWISLISYANTMKTSSKKLIMINKENTE--	2166
Qy	1111	IKCGDLFE--NSVVDFFNECAVSRKCVPRKSDLG-EFPAPDPSPVLQVNFNISDFNGKWY	1167
Db	2167	-KVDYLIKDNSSSTDGYVETL-----KGYGSKLTFSSASEIVONADTYSVNFPAKH	2216
Qy	1168	ITSGLNPTFDADFQCLGHEHTEGDNKLVGNISWRIKTLDGFFTRSAVQKVFQDPNPGV	1227
Db	2217	EKESLNAIRD- IKELLYLFHQNSDISIV-----EGGVQNMNL-----	2252
Qy	1228	LYNHNDYLYHQDDWLTLSKSIENKPEDY-----IFVYVRGRNDA-----	1267
Db	2253	LYDKLNEEKREMDLYRNISETKLQOMESHSTDFVKPMIELHKGMMETNNKSLEKEKKLK	2312
Qy	1268	-----WDYGGAVVYTRSSV--LPN--SIPELEKAAKSIGRD-----	1301
Db	2313	SVNDHMSMEAEMIKNGLKYTPESVQNIINIIYSVTEAEVKTLEEDRIDYGDYQVIEEHK	2372
Qy	1302	--FSTFTRTDNTCGPEPALVERTEKVERGE-----RIIVKEVEETEEVEERKEVEKV--GR	1353
Db	2373	KQFSILLDRTN-----ALMDDEIEFKENNYNLMVEVTEIHRVNDYIEKITNKLQVAK	2426
Qy	1354	TE-----MTLQR---LARGFNELKQDENFVRELISKEEM-----EFIDE	1390
Db	2427	TEYEQILENTIKQNDMLQNIIFLKVKVSIIEFFYFNVKKKKESILINDLYEQERLLKIGHLDE	2486
Qy	1391	IKMEASE	1397
Db	2487	IKRNVTE	2493

RESULT 10

H36812

hypothetical protein ORF64 - saimirine herpesvirus 1 (strain 11)

A:Species: saimirine herpesvirus 1

A:Note: host Saimiri sciureus (common squirrel monkey)

C:Date: 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change 08-Oct-1999

C:Accession: H36812

R:Albrecht, J.

submitted to the EMBL Data Library, January 1992

A:Description: Primary structure of the herpesvirus saimiri genome.

A:Reference number: A36806

A:Accession: H36812

A:Molecule type: DNA

A:Residues: 1-2469 <ALB>

A:Cross-references: GB:X63436; NID:g60320; PIDN:CAA45687.1; PID:g60385

R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; N

J. Virol. 66, 5047-5058, 1992

A:Title: Primary structure of the herpesvirus saimiri genome.

A:Reference number: A37309; MUID:92333688

A:Contents: annotation; protein-coding frames

A:Note: neither protein nor nucleotide sequence is given

C:Genetics:

A:Gene: 64

RESULT 10

H36812
 hypothetical protein ORF64 - saimirine herpesvirus 1 (strain 11)
 C:Species: saimirine herpesvirus 1
 C:Note: host Saimiri sciureus (common squirrel monkey)
 C:date: 16-Oct-1992 #sequence_revision 16-Oct-1992 #text_change 08-Oct-1999
 C:Accession: H36812
 R:Albrecht, J.
 submitted to the EMBL Data Library, January 1992
 A:Description: Primary structure of the herpesvirus saimiri genome.
 A:Reference number: A36806
 A:Accession: H36812
 A:Molecule type: DNA
 A:Residues: 1-2469 <ALB>
 A:Cross-references: GB:X64346; MID:g60320; PIDN:CAA45587.1; PID:g60385
 R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.;
 J. Virol. 66, 5047-5058, 1992
 A:Title: Primary structure of the herpesvirus saimiri genome.
 A:Reference number: A37309; MUID:92333688
 A:Contents: annotation; protein-coding frames
 A:Note: neither protein nor nucleotide sequence is given
 C:Genetics:
 A:Gene: 64

Query Match 2.1%; Score 155; DB 2; Length 2469;
Best Local Similarity 17.3%; Pred. No. 1.2;
Matches 230; Conservative 208; Mismatches 497; Indels 39

QY 276 VOTFVQDPLGALYNHNEFLHYQ-----DDWYILSSQIENKPPDDYIFVYIRG 324
Db 949 YE-----DLVKCKENDMEFLKESAKLGHVTVSNEAYSELEKKLEQPSLEYLVEHAKA 1002
QY 325 RND---AWDGYGGSV-----IYRSPITPESIIPN-----LOKAASVGRDF--- 363
Db 1003 TNHLLSDSAYEDLVCKENPDMEFLKESAKLGHVTVSNEAYSELEKKLEQPSLEYLVE 1062
QY 364 -----NNFITDNS-----CGPEPPLVERLEKTAABEGEKLIIKEAV-ELEEVEKE--- 408
Db 1063 HAKATNHLLSDSAYELVCKENPDMEFLKESAKLGHVTVSNEAYSELEKKLEQPSLEY 1122
QY 409 --VEKVRDTEMTLFORLEGFKELOQDEENFVRELSKEEKEILNE--LOMEA-TEVEKLF 463
Db 1123 YLVHAKATNHLLSD--SAYEELVCKENPDMEFLKESAKLGHVTVSNEAYSELEK-- 1178
QY 464 GRALPIRKLKALAPHNSFLANHETIKYVGVSKLPGHKRFSWGWEDYFGSIVVAK----- 518
Db 1179 -----KLEQ---PSLAYLVEH-----AKATDHLLS---DSAYEDLVCKENPDV 1217
QY 519 -----ICSSRRIPRYFRKSPRICCGLDLSRGL-QLFSGHKNLSPAHNSINQNV 564
Db 1218 EFLKESAKLGHVTVSNEAYSELEK-----LEQPSLAYLVEHAK--ATDHLLSDSA 1268
QY 565 PKNSGCKFPKDVAVLWVEKWGQPAKTAIVAFITLSVASKAD-----AVDALKTCTCL 617
Db 1269 YEDLVCKENPDMEFLK-ERKSAKLGHTVVSNEAYSELEKKLEQPSLEYLVEHAKATNHLL 1327
QY 618 LKECRLE-LAKCISNPAC-----AANVACLOTNNRPDTEQCIKCGDLFENSVDDEFN 670
Db 1328 LSDSAYEDLVCKENPDMEFLKESAKLGHVTVSNEAYSELEKKLEQPSL----- 1377
QY 671 ECAYSRRKCVPRK--SDVGDF-----PVPDPSVLVQKFDKDFSGKWFITRGLNPTFD 723
Db 1378 EYLKHAQIOQSKIIISDFNTLANPSMEDMAKLOKLEYQIVSNDEYI--ALKNTMEKP 1435
QY 724 DCQ-----LHBFH-----TEENKLVGNLSWRIRTPDGGFFTRSAVQKVEQD-PKYPGILYN 773
Db 1436 DVELLSKLGHYHIIDTTYNELVSNFN-----SPTLKFIEBKAKSGYRLI 1482
QY 774 HDNEVL-----LYODDWYILSSK-----VENSPE-----D 798
Db 1483 EPNEVLNRLNRTTPSKKEIDNFCQIGCVALDSKEYERLKNLSLENPKFIEBENAALLD 1542
QY 799 YIFV---YKGRNDWDGYSGLYVTRSAVLESIIPELOTAQKVGDRDNTFTKTDNTC 855
Db 1543 LVLVDKTEYQAMK-----NASNKSILPSTRAL-----DFVTM----- 1576
QY 856 GPEPPLVERLEKKVEEGERTIIKEVEEIEEVEKVRDKEVTLFSKLEGEKELORD--- 911
Db 1577 -PAPQASAEEKSLQK-----RTLSDIENELKAL-----GYVAIRKENLPN 1616
QY 912 -EENFLRELSKEEM-----DVLDDGLKMEATEVEKLF-----RALPIR 948
Db 1617 LEKPIVDNASKNDVLNLCSEKSLVPLSTEEYDNNRKEHTKILNILDGDSIDFLKECEKY 1676
QY 949 KLMAYATH-----CFTSPCHDRIRFFSDGIGRLGIRTKRINGTFLKILPDIQSAD 1001
Db 1677 QMLIISKHDYEEKQBAIENPGYEFLKESALGY-----ELVSEVELDRMKQIDSPD 1729
QY 1002 LRTTGRRSRPLSAFRSGFSKGIPIVDPLPSKNELKELTAPLLKLVGLACAFILVPSA 1061
Db 1730 IDYMOEKAARN-----EMVL--RNEEKE-----ALOKKIEYSLTFLIEKAA 1770
QY 1062 -----DAVDALKTKACLLKGCRIELAKCIANPACAAVACLOTNNRPDTEQCIKCG 1114
Db 1771 GNNKILVDQIEYDEI-----IRKC-----NHPTRMEELESCH 1802
QY 1115 DLFENSVDDEFNECAVSRKCKVPRK--SDLGEFPAPDPSVLVQNVFNISDFNGKWTITSGLN 1173
Db 1803 HL--NLVLDDQNEYSTLREPLENRNVEDLINTLSKLYAIAPNTIYQDLIGKYE-----N 1855
QY 1174 PTFDADFQCLHFEHTEGDKLVGNISWRIKTLDLSGFFTRSAVQKVEQDPPNPGVLY---- 1229

Db 1856 PNFDYVKDSLANK-----MDYVAISRQDYELMWAKYKQPOLDYILKIS 1896
QY 1230 -----NHDNEYLVHQQDDWYILSSKIENKPEDYIFVYVGRNDWDG 1270
Db 1897 SKDHIHVPLSEYNLMVTNRPNSLSYLKKAVALNNHILLIKEDDYKNIL-----AVSE 1950
QY 1271 YGGAVVYTRSSPLNSIILEPELEKAAKSGISGRDFTFIRTDNCTGPEPALVERIEKTVEEGE 1330
Db 1951 HPTVTHLSEKASLLKNVLVDKD-----DFATMSRSIE-----KPTIDFLSTKALSMG- 1997
QY 1331 RIIVKEV-----EETEEVEKEVEKVGRTMTLFORLAEGFENELKQDEENFVRELSKEEM 1385
Db 1998 KILVNESHTRKNEKILLSEPDSEFTMKAKEQGLIIISEKEYSELQDIDRPSLDVLKEKA 2057
QY 1386 EFLDEIKMEASEVEKLFKALP 1407
Db 2058 AIFDSIIIVENIEYQOLVNTTSP 2079

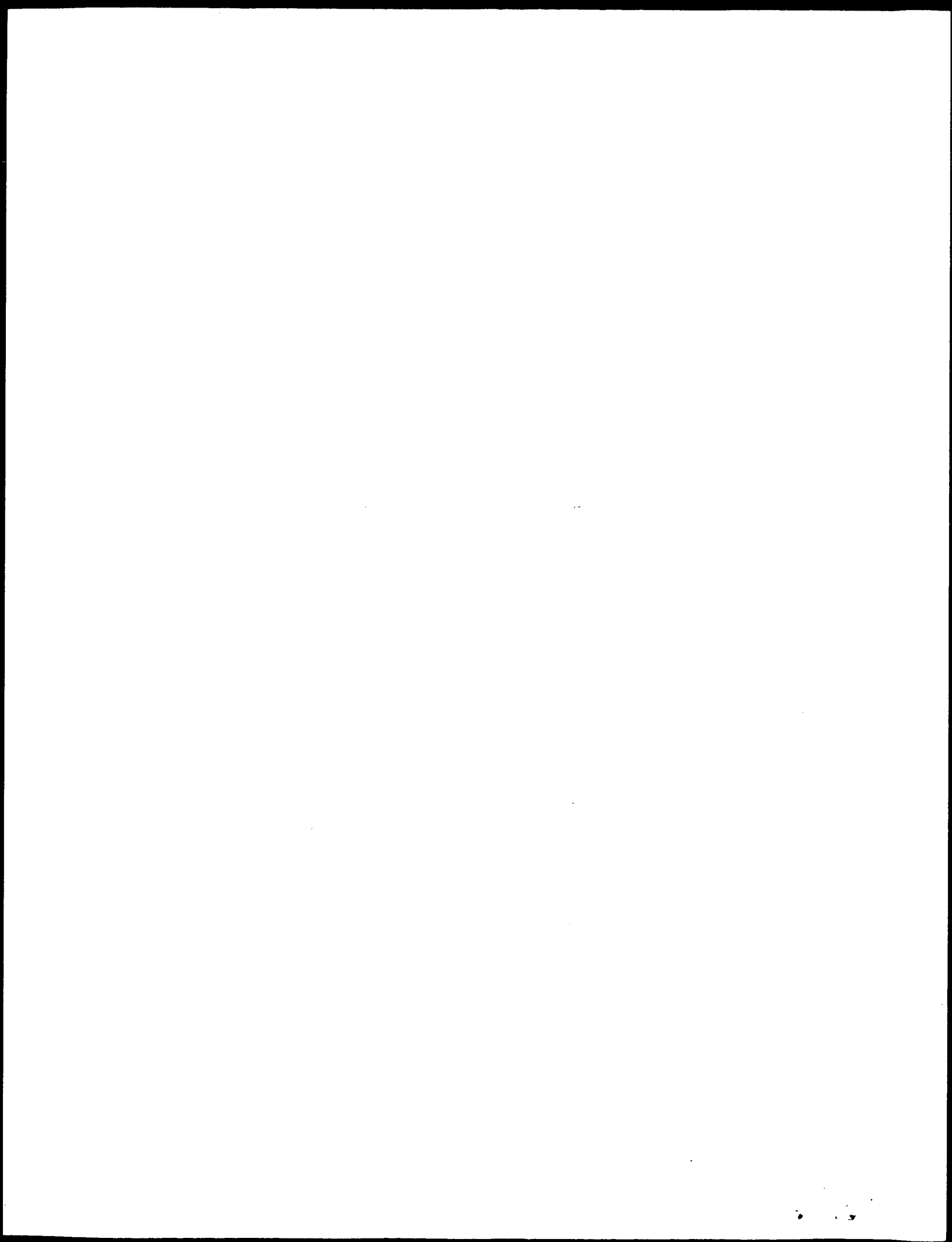
RESULT 15
S54174
DNA topoisomerase (EC 5.99.1.2) - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jun-2000
C:Accession: S54174
R:Tosh, K.T.; Kilbey, B.J.K.
A:Description: Isolation and characterisation of the topoisomerase 1 gene from Plasmodi
A:Reference number: S54174
A:Accession: S54174
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <TOS>
A:Cross-references: EMBL:X83758; NID:g790481; PID:g790482
C:Superfamily: eukaryotic type I DNA topoisomerase
C:Keywords: DNA binding; DNA replication; isomerase

Query Match 1.9%; Score 144.5; DB 2; Length 839;
Best Local Similarity 17.3%; Pred. NO. 1.1;
Matches 191; Conservative 176; Mismatches 390; Indels 347; Gaps 46;

QY 352 LOKAASVGRDNFNFTTDSGPEPPLVERLEKTAABEGEKLIIKEAVEI---EVEEVEK 409
Db 24 INKIKONTLG---NN--KSCNSRSSKRESIKKOKSSELGKIKNTKSKLGKKEEKKQI 78
QY 410 EKVRDTEMTLFORLEGFKELOQDEENFVRELSKEEKEILNELOMEATEVEKLFGRALPI 469
Db 79 SRKSNELKEKNLKEGKKYVEKKSRV---KDETILTNVIKKETQNNKK-----PK 128
QY 470 RKLRLMALAPHSFLANHETIKYVGVSKLPGHKRFSWGWEDYFGSIVVAKICSSRRIPRYF 529
Db 129 KLLKKS---EENF---EPINRW-WEKIDQDIDIOWNY----- 158
QY 530 RKSPRICCGLDLSRGLQFSGHKNLSPAHNSINQNVPKNSGCKFPKDVAVLWVEKWGQFA 589
Db 159 -----LEHURL-IF-----SPPY-VQHHVP-----IFYSIKIEL-----NA 188
QY 590 KTAIVAIFILSVASKADAVDAKTCCLKECRLEIAKISNPACAAVACLOTNNRPD 649
Db 189 KSEELATYWC-----AIGSDYCTKEKFLNFKTFIN-----SLENDNIQK 231
QY 650 EYECQIKCGDLFENSVDDEFNECAVSRKCKVPRKSDVGDFPVPDPSVLVQKFDM----- 703
Db 232 ENTKLKKGDI SNFKPID-----FMPKIDHLLKREELKNTKEE 271
QY 704 KDFSGKWFITRGLNPTFDADFQCLHFEHTEENKLVGNLSWRIRTPDGG-----FFTRSAV 758
Db 272 KEKEKRWMEKELPYTYVALVDWIREKISSNKAEPPLGFRGEGHPKQGLKKRIFPEDV 331
QY 759 QKVPQPKYPGILYN-----HDNE--YLLYODDWYILSSKVENSPEDYIFVYK 806

Db 332 INISKDAPVPRLYDNMCGHNWDIYHDKVWTWLAYKD-----SINDQIKYTFLS 381
Qy 807 RNDWDGYSVLYTRSAVLPESEIPELQTAQKVGKRDFTFKT-----DNTCGPEPPLV 862
Db 382 AQSKFKGKDKLMKYENAR-----KLKSCVHKIREDYKMKMKNIIDKQLGTAVYLI 433
Qy 863 ERLEKKVEEGERTTIKEVEETEEREEVEKVRKQEVTLFSLFEGFKELQORDEENFLRELSKE 922
Db 434 DFLALRV-GGEK-----DIDEADTV-----GCCSLRVEHISFAHDIPFK 472
Qy 923 EMDVLDGLKMEATEVEKLFGRALPIRLKMAVATHCFTS---PCHDRIRFFSSDDGIGRLG 979
Db 473 SVD-----SKEQKTNDKVNKTPLENTLESISSEDCYITLDFLGKDSIRYENT----- 520
Qy 980 ITRKRINTFLKLPPIQSDADLTGTGRRSRPLSAPRSFGSKGIFDIVPLPSKNE-LKE 1038
Db 521 ---VKIDKQAYINII-----IFCKNKNRDEGVDFQITCSKLINEYLKE 559
Qy 1039 LTAPILLKLVGLACAFILVPSADAVDAL--KTCACILLKGCRIELAKCIANPACAAVAC 1096
Db 560 IMPTLSAKVFTYNNASITLDOOLKRIKEVYKTKTYSLSG-ETELHK-----SK 607
Qy 1097 LQTCNNRPDETCQIKCGDLFENSVDDEFNECAVSRK----KCVPRKSDLGEPAPDPSPV 1152
Db 608 KRKSSHLTSDTNILSDASDSTINDVNNEYDENGINKLSYATTYVGKENDVDDKNSP-IEV 666
Qy 1153 LVQENISDFNGKWYITSGLNPTFDQFCQLHEFHTEGDNKLVGNISWRIKTLDGFFTR 1212
Db 667 DVSNN-----ELINFYNNANREVAIILCNHORSIPKQHDITM 703
Qy 1213 SAVOKFVODPNQPGVLYNHD-NEYLHYQDDWYILSSKIENKPEDYIFVYVYGRNDAMDGY 1271
Db 704 SKIKKQIE-----LYNEDIKEYKY-----LQHLKKNSDKKFIFV----- 738
Qy 1272 GGAVVYTRSSVLPNSIIPLEKAASIGRDFSTFIRDTNCGPEPALVERIEKTVEEGER 1331
Db 739 -----SKVSTLDGTLRP-----NKKVENMK 758
Qy 1332 IIVKEVEIEEVEKEVEKVGRTENTLQRLAEGFNLKQDEENFVRELKSEMEFLDE- 1390
Db 759 -----EESCKKKL-----ITLIKVELLNQMKVRDNDKNTALGTSKINYMDPR 802
Qy 1391 -----IKMEASEVEKLFCKALPIR 1409
Db 803 ITVAFCKKFEPIEKVFNRLK 826

Search completed: November 6, 2001, 05:03:18
Job time: 6742 sec



Result No.	Query			ID	Description
	Score	Match	Length		
1	158.5	2.1	1957	1	YD86_SCHPO
2	155.5	2.1	2869	1	R9P1_PLAYBA
3	155	2.1	2469	1	TRGU_HSVSA
4	148.5	2.0	1935	1	MSB_PIG
5	146	1.9	1935	1	MSB_HUMAN
6	143	1.9	2748	1	N1M1_YEAST
7	142.5	1.9	1935	1	MSB_RAT
8	139.5	1.9	1025	1	M2P1_YEAST
9	137.5	1.8	1875	1	MLP1_YEAST
10	135	1.8	3660	1	DMD_CHICK
11	134.5	1.8	1937	1	M2SP_HUMAN
12	134	1.8	1630	1	MSPI_PLAYBK
13	134	1.8	1639	1	MSPI_PLAFA
14	133.5	1.8	1934	1	MSB_MESAU
15	132	1.8	809	1	LEF_BACAN
16	130	1.7	1679	1	Y1O9_YEAST
17	129.5	1.7	1805	1	HMW2_MYCE
18	129	1.7	1822	1	YM68_CAEEL
19	128	1.7	2230	1	GOG4_HUMAN
20	127.5	1.7	1790	1	USO1_YEAST
21	126.5	1.7	4568	1	DYHB_CHLRE
22	126	1.7	993	1	TSB_LACIA
23	126	1.7	1744	1	TANA_XENLA
24	125.5	1.7	838	1	HIS2_CANAL
25	124	1.7	2284	1	POLI_GFIV
26	123.5	1.6	690	1	SYGB_BUCAI
27	123.5	1.6	761	1	METE_AQUAE
28	123	1.6	3678	1	DMD_MOUSE
29	122.5	1.6	1341	1	REAL_METJA
30	121.5	1.6	1039	1	SVI_METJA
31	121.5	1.6	1170	1	TSPI_BOVIN
32	121.5	1.6	1545	1	YMH2_YEAST
33	121	1.6	2025	1	TTC3_HUMAN

```
QY 337 IYTRSPITPSIIIPNLOKAQSVGRDNFPIITDSCGPEPPPLVERLEKTAEGEKLLIK 396
Db 545 -----ISNHELSLSQSLTAAEKAQAVATNNE-----LSGKNSLQT 581
QY 397 EAVEIEEVEKEVKVDTDM-----TLFORLLEGFKELQOQDEENFVRELSKEKEILN 450
Db 582 LCNAFOQLAKSVNQLKENEQNFSSLDTSFKLNESHQLENNHQITQOL-KDTSKQLQ 640
QY 451 ELQMEATEVEK-----LFRGALPIRKLRMALAPH-----SNFLANHETIKYIVGS 495
Db 641 QLQLERANFEQKSTLSDENNDLRTKLLKEESNKSLLIKQEDVDVSLKXNIQTLK----- 695
QY 496 KLPQHKRFSWGWDYFGSIVVAKICSRRIIPRYFRKSPRICCGLDGRGLQF-----SHGKH 552
Db 696 -----ED-----LRKSEALRFSK-----LEAKNREVINDNLKGX 726
QY 553 NLSPA-----HSI-----NONVPKNGSGCKFPKPDVALMVWEKQGQAKTAIVAFILSVA 602
Db 727 ETLAQRNDLHSSDAKNTNAILSELTKSSDV-----KRUTANVETLTQD 774
QY 603 SKA-----DAVDALKTCCTCLLKECLELAKCISNPACAAVACLOTCNNRPDETECQIK 656
Db 775 SKAMKQSTSLVNSYQSISNLYHELDRDHYN-----MQSQNNLTLESSEK 821
QY 657 --CGDLFEN--SVVDEP-----NECAVSRKKCVPRK-----SDVGDFPVPDP 694
Db 822 TDCENLTQNNLTLDNVQKLMHKHVNOESKVSSELKEVNGKLSLDLKNLRSNVAISDND 881
QY 695 SVLVOKFDM-KDFSGKWFITRGLNPTDFDQCQLHEFTEE-----NKLGNLSWRIR 746
Db 882 QILQLAELSKNYSLSQESQAUNGLSKLSLEAEKQLLHTENEELHRLDKLTGKLK----- 937
QY 747 TPDGGFFTRSAVQVQDPYPGYGLYIYHNDNEYLLODQWYILSKSVENSPEDIFVYIKG 806
Db 938 -----IEESKSSDL-----GKLLTARQEE--ISNLKENMSQQAITSVKS 976
QY 807 RNDWDGYSGLVTRSAVLPESIIPELOTAQKVGDRFNTFKTDNTCQPEPLVERLE 866
Db 977 KLDE-----TLSSKSKL-EADIEHLKNKVSSEVERNALLASN-----ERLM 1017
QY 867 KKVE-EGERTIIKEVEIEEVEKVRQKVTFLSKL-----PEGF-----KELQD 911
Db 1018 DDLKNGE-----NASIQTEIEKRAENDDLQSKLSVSVSEYENLLISSQTNKSLD 1071
QY 912 BENFLRELSKEEMVDVGLKMEATEVEKL-----FGR-----ALPIRKLMAVATHC 957
Db 1072 KTNQLKYTEKNVQKLLDEKQDNVLEBELTSKYGLGEENAQIKDELLALRKK----- 1124
QY 958 FTSFCHDRIRFFSSD-----DGIGRL-----GITRKRINGTFLKILPPIQS--AD 1001
Db 1125 -SKQOHLCAHFVDOLKEKSDALEQNLNEXNELIVSLEQNSNNEALVEERSDLANKLSD 1183
QY 1002 LRTTGGSSRPLSAFRSGFSKGIDFIVPLPSKNELKELTAPLLIKLVGLACAFIIVPSA 1061
Db 1184 MKSLSDSNDNISVIRSD-----LVRVN 1206
QY 1062 DAVDALKTACLLKGRTELAKTANPACAAVACLOTCNNRPDETECQIKCGDLFENS 1121
Db 1207 DELDTLK-----KDKSLSTQYSEVCQDRDLDLDSLKGC----- 1240
QY 1122 VDENECAVS-RKKCVPRKSDLGFPAPDPVIVON--FNISDPNGKWYT-SGLNPTD 1177
Db 1241 EESNKYAVSURELCTSEIDV-----PVSEILDNDNFVFNAGNFSELSRUTVLSLENYLD 1295
QY 1178 AFDQQLHEFHTEGDKLVGNISWRIKTLDSGFFTRSAVQKRVQDPNPGVLYNHDNEYLH 1237
Db 1296 AFN-QVNFKKMELDN-----RLTTDAEF-----TKVVDL-----EKLOH 1330
QY 1238 YODDWYILSSKIENKPEDYIFVYIRGRDNDWDGYGAVVYTRSSVLNSIPELEKAAS 1297
Db 1331 EHDWLI-----ORG-----DLEKALD 1348
```

```
QY 1298 ICRDFSTFRTDNTCGPEPALVERIEKTVGERLIIVKEVEIEEVE----- 1345
Db 1349 SEKNP-----LRKEAENTENH-SLEGKEETKKEIAELSSRLDNQLATNKLKNQ 1398
QY 1346 -----KEY-----EKVGRTEMTLFQRLAEGFNLKODENFVRELSKEEMEFL--DEIKMEA 1395
Db 1399 LDHNLQNEIRKEDVLKESLIISLEESLSNQROKESLLD--AKNELEHLDLDDTSRNS 1456
QY 1396 SEVEKL 1401
Db 1457 SLMEKI 1462

RESULT 2
RBPL_PLAVB
ID RBPL_PLAVB STANDARD; PRT; 2869 AA.
AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN RBPL.
OS Plasmodium vivax (strain Belen).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_taxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites."
RL Cell 69:1213-1226(1992).
CC -!- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -!- SUBUNIT: HOMODIMER (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M80997; AAA29743.1; -.
DR HSP: P36956; IAM9.
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
FT SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;

Query Match 2.1%; Score 155.5; DB 1; Length 2869;
Best Local Similarity 17.4%; Pred. No. 0.86;
Matches 283; Conservative 242; Mismatches 573; Indels 529; Gaps 76;

QY 57 TSYKTSFSDSSSHCKDQKQICSDITSFEEI-----QRFDLKRGMTLILEKQWRQFIQLA 110
Db 1150 TSNKNEILKSVKEVEDKLNLEQNEDEYKVKPENEKQLEAIRGSMKLE----- 1200
QY 111 IVLVCTFIVIPRVDAVDALKTCACLLKECRIELAKCIANPACAAVACLOTCNNRPDETE 170
Db 1201 -----VINKHVSMTQLESTANTLK-----SNAGKENEHDEELN----RKT 1239
QY 171 CQIKCGDLFE--NSVVDQFNECAVSRKKCVPRKSDVGEPFVDRNA--VVQNFNM-KDF 224
Db 1240 GQMR--DIYEKLLKIAELKEGTGVNELKDANEKANKE--PEPERNIIGHLEKITVENDK 1296
```


QY	1111	IKGCDLFE--NSVVDFENECASRRKKCVPRKSDLG-EFPADDPVSLVQNFNISDFNCKWY	1116
Db	2207	-KVDYIKONSSDGDGYETL-----KGFYGSKLTFSSASIVQADTYSVNFQKH	2256
QY	1168	ITSGLNPFTDAFCQCLHFEHTEGDKNLGVNISWRIKTILDSGFFTSRVAQKVFQDPNPGV	1227
Db	2257	EKESLNAIRD-IKKELYLFHQNSDISV-----EGGVQNM-----A	2292
QY	1228	LYNHDEYLYHQDDWYILSKSIENKPEDY-----IFVYVRGRNDA-----	1267
Db	2293	LYDLKNAEKEKREMDELYRNISSETKLQMEHSTDFVFKPMIELHKGMMNETNKKSLLEKEKKLK	2352
QY	1268	-----WDQYGCAGVYTRSSV--LPN--SIIPLEKKAASKTGRD-----	1301
Db	2353	SVNDHHSMEAEKMKGLKYTPESVQNNINNIYSVTEAEVKLEIDRDYGDNYQIVBEHK	2412
QY	1302	--FSTFRTDNTCGPPALVERIEKTVERGE-----RIIVKEVEEIEEVEKEVKV--GR	1353
Db	2413	KQFSILIDRTN-----ALMDMIEFKKENNYNLMEVNTETIHRVNDVIEKTNKLVQAK	2466
QY	1354	TE-----MTLFOR--LAEGFNEQLQDRENFVRELSKEEM-----EFLDE	1390
Db	2467	TEYEQILENIKQNDMLQNFILKVKYSIIIEYFENVKKKKESILNDLYEQERLLKLTGEHLDE	2526
QY	1391	IKMEASE 1397	
Db	2527	IKRNVT 2533	
RESULT 3			
TEGU_HSVSA			
ID	TEGU_HSVSA	STANDARD;	PRT; 2469 AA.
AC	Q01056;		
DT	01-APR-1993 (Rel. 25, Created)		
DT	01-APR-1993 (Rel. 25, Last sequence update)		
DT	01-APR-1993 (Rel. 25, Last annotation update)		
GN	PROBABLE LARGE TEGUMENT PROTEIN.		
DE	64 OR EEF2.		
OS	Herpesvirus saimiri (strain 11).		
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
OC	Gammapherpesvirinae; Rhadinovirus.		
OX	NCBI_TaxID=10383;		
RN	[1]		
RX	SEQUENCE FROM N.A.		
RP	MEDLINE=92230228; PubMed=1314457;		
RA	Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,		
RA	Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,		
RA	Honess R.W.;		
RT	"Primary structure of the herpesvirus saimiri genome.";		
RT	J. Virol. 66:5047-5058(1992).		
RN	[2]		
RX	SEQUENCE FROM N.A.		
RX	MEDLINE=92230228; PubMed=1314457;		
RT	Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;		
RT	"Analysis of nucleotide sequence of the rightmost 43 kbp of		
RT	herpesvirus saimiri (HVS) L-DNA: general conservation of genetic		
RT	organization between HVS and Epstein-Barr virus.";		
RL	Virolgy 188:296-310(1992).		
CC	!- FUNCTION: TEGUMENT PROTEIN.		
CC	!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,		
CC	EHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-		
CC	tion between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	at the European Bioinformatics Institute. There are no restrictions on		
CC	use by non-profit institutions as long as its content is in no		
CC	modified and this statement is not removed. Usage by and for commerc-		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X64346; CAA5687.1; -;		
DR	EMBL; M86409; AAA46140.1; -;		
DR	PIR; H36812; H36812.		

SQ SEQUENCE 2469 AA; 280165 MW; D2B4B8DC08644CDB CRC64;
 Query Match 2.1%; Score 155; DB 1; Length 2469;
 Best Local Similarity 17.3%; Pred. No. 0.75;
 Matches 230; Conservative 208; Mismatches 497; Indels 396; Gaps 58;

Qy 271 FTRSAVQTFVDDPDLPCALYNHDFLHYQDDWYLLSSQIENKPDYIFVYR-----323
 Db 178 FSTSNVDAIEYLPNNVQYT--GSFLYFVPEYVIGHSH-----YIMNHRYVINYEKL 228
 Qy 324 -GRNDAMWGYGSGVIY-----TRSPUPESIINLOKAAKSVGRDNFNI 367
 Db 229 HGNPIDLTSQELLIEISPPNTPKPTSTQPKPTPTPKPATKPKTPKPKTPRESTI 288
 Qy 368 TTDNSCGPEPLPLRLKTAABEGEKLIIKAV-----EIEEVEKEVEKVRDT--EMTL 419
 Db 289 PYDKS--KKPP--KIPKTSKKKKVLTQDALTPTQHKTIIEHLRELLPPTITVEDNTL 343
 Qy 420 FQR-----LLSGFKELQ-----QDENFVRELSEKKEIILNLOMEATEVEKL 462
 Db 344 FNPVVERTPTGDSLSGINSSTTKREDLDDDDNVTSKLKEDDDWIDD-----392
 Qy 463 FGRALPIKRLMALAPHSFLANHETIKYVGSKLPGHKRFSWGEDYFGSIYVA-----517
 Db 393 ---IPIPEVLDTEHSH---DOETI-YMGDE--NIHDSYSDDDIDDTLDSFIQLD 441
 Qy 518 -KICSSRRIPRYFKSPRICGLDSRGLQFHSOK-----HNLSPAHSINQNPVK 567
 Db 442 NLITSLOINPKN--NTFPRI---IDKTSNQPTKEGKALHSIDRLKNTLVLEHGLTSSIS 497
 Qy 568 NSGCKFPKDALWKEWKGQFAKTAIVAFITLSVASKADAVDAKLTCTCLIKECRELEAK 627
 Db 498 ISKCKSLQFVIL---WGE-----KLSIPTR-DLKTLIKTELLITEAETALTK 542
 Qy 628 CISNPACAAANVACLQTCNNRPDETECOIKGCDLPENSVDDEFNECAVSRKKVCPRKSDVG 687
 Db 543 -LTN-----DTPRNVITKLNCKMLKLK-----SESVD 569
 Qy 688 DFP-----VPDPSVLVQKDM-----KDFS-----GKWFITRGLNPTF 720
 Db 570 SYKHLALLNIIKIOTIDIEIELKLTSTVFTSELGKDFSVVCTKKESETIMAAIKNLK 629
 Qy 721 DAFDCOLHEPTEENKLVNLSWRIRTPDGGFFTRSAVQKF-----VQDKPYG 769
 Db 630 EKISTRKQELHEENYFOSVL-----IAMEFQPIPLTRVIEIQPSKKAQ 675
 Qy 770 ILYNHDEYLLYQDDWYTLSSKVENSPEDYIFVYKGRND---AWDGGGSVLYTRSAVL 826
 Db 676 QL--HEKSKLVEQK---LTIDANNVLTDLHTMKQDKTDISPAPD-----FTTVLKN 722
 Qy 827 PESIPELOTAQKVRD---FNTFIKTDNTCGPEPLVLEKKEVEGERTI-IKEVEE 882
 Db 723 IQSTIQLLQTCVTDLNIDKDKFISNTVQQLSYIGWEVAELSHSNWFFPKADPVIPKLITDD 782
 Qy 883 IEEVEKVRDKEVTLFSKLFEGFELQDENFRELSEKEMDVLGDKWEATEVEKLFG 942
 Db 783 IKKELOQVTTKQ---KNETLSKILADVOTLLENKQSDTSLIPILQHYIYKAGTLVG 837
 Qy 943 RALPIRKLMAVATHCTSPCHDIRFRTSSDDGIGRLGIT-----KRRINGFTLLKILPPTQ 998
 Db 838 E-----RENOKFESLKNVQKLSSTSEFLTKLIDSTLLENVQLQIQ 878
 Qy 999 S-ADLRTTGGSSRPLSAFRSGSKG---IPDIYPLSKNELKELTAPLLL-----1045
 Db 879 EISDILQSNQYIHOSETIKQAFFDKRSNTIINNLIQINQOQYTTVTQPMIAVKARFUSEA 938
 Qy 1046 -----KLVGLACAFILVPSADAVDAKLTCAKCIKKGRIELAKCIANPACAAANVAC 1096
 Db 939 KPRESNTICEIISTLVLSGLSKSTTVTEALKDALKSIDTLKEL-----983
 Qy 1097 LOTCNRNRPDETE-----COIKGDDLPENSVDDEFNECAVSRKKVCPRKSDLGEFPAPD 1149

Db 984 --TAVDRPLKRELYNVIRKLOKQLKTLLEQQEFDNW-----KMEVDSF-VPT 1027
 Qy 1150 PSVLVQNF--NISDFNGKWIYTSGLNFTDFADFCOLHEFHETEGDN-KLVGNISRIKTLTD 1206
 Db 1028 PSRDYKTTIQNAPSMMKAKOYAKKALKDQIQAMEIDVDPESVIEDNIKANGOKAQW-----1082
 Qy 1207 SGFFTRSAVQKRVQPNQPGVLYNHNHDEYLLHYQDDWYTLSSKIKENKPKDYIF-----VY 1261
 Db 1083 -----KIQSAFDLNFSLI-----PDDWLSL-AKEYTRPKSTLFTVIGPILL 1124
 Qy 1262 RGRNDAMWGYGAVVYTRSSVLPNSII-----PELEKAAK 1296
 Db 1125 KEVEEVLESVKNLKEAKLSLLPNGPVFTPPKFDWIHYHESNVNPFHLKTLNLPKVSIVAH 1184
 Qy 1297 SIGRDFSFIRTDNT-CGPEPALVERIE-----KTVE-----EGERIIVKEV 1337
 Db 1185 NIGHELSSLLSQALNSKTLPEAVVGTSLSEQHAAKFCMFKTLEATWHDHVDVTRTKIDYI 1244
 Qy 1338 EETEEVEK-----EVEKVGVR-----TEMTLFQRLAEGFNELKQDENFVRELSEKEMEFL 1388
 Db 1245 EDLRNDTKKHIVAPQIQSPNRFSLPEDIQEIINSLPKLFRDSLLENESRLASQKNEFQML 1304
 Qy 1389 DEIKWEASEVE 1399
 Db 1305 -ENTVKAABLQ 1314
 RESULT 4
 ID MYSB_PIG STANDARD; PRT: 1935 AA.
 AC P79293;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
 GN MYH7.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=DOMESTICA;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS OF THE MYOFIBRILS.
 CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
 CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE..
 CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U75316; AAB37320.1; -
 CC InterPro: IPR000048; -

MEDLINE=95179132; PubMed=7874131;
 RA Greve G., Bachinski L., Friedman D.L., Czernuzewicz G., Anan R.,
 RA Towbin J., Seidman C.E., Roberts R.;
 RT "Isolation of a de novo mutant myocardial beta MHC protein in a
 RT pedigree with hypertrophic cardiomyopathy.";
 RL Hum. Mol. Genet. 3:2073-2075(1994).
 RN [22]
 RN VARIANTS CMH1 CYS-513; ARG-716, AND TRP-719.
 RX MEDLINE=94110336; PubMed=8282798;
 RA Anan R., Greve G., Thierfelder L., Watkins H., McKenna W., Solomon S.,
 RA Vecchio C., Shono H., Nakao S., Tanaka H., Mares A. Jr., Towbin J.A.,
 RA Spirito P., Roberts R., Seidman J.G., Seidman C.E.;
 RT "Prognostic implications of novel beta cardiac myosin heavy chain gene
 RT mutations that cause familial hypertrophic cardiomyopathy.";
 RL J. Clin. Invest. 93:280-285(1994).
 RN [23]
 RN VARIANTS CMH1 THR-797.
 RX MEDLINE=96047159; PubMed=7581410;
 RA Woolman J.C., Brink P.A., Corfield V.A.;
 RT "Identification of a novel Ala797Thr mutation in exon 21 of the beta-
 RT myosin heavy chain gene in hypertrophic cardiomyopathy.";
 RL Hum. Mutat. 6:197-198(1995).
 RN [24]
 RN VARIANTS CMH1 CYS-453.
 RX MEDLINE=96209901; PubMed=86551135;
 RA Ko Y.-L., Chen J.-J., Tang T.-K., Cheng J.-J., Lin S.-Y., Liou Y.-C.,
 RA Kuan P., Wu C.-W., Lien W.-P., Liew C.-C.;
 RT "Malignant familial hypertrophic cardiomyopathy in a family with a
 RT 453Arg-->Cys mutation in the beta-myosin heavy chain gene:
 RT coexistence of sudden death and end-stage heart failure.";
 RL Hum. Genet. 97:585-590(1996).
 RN [25]
 RN VARIANTS CMH1 THR-349 AND TRP-719.
 RX MEDLINE=98204402; PubMed=9544842;
 RA Jeschke B., Uhl K., Weist B., Schroder D., Meitinger T.,
 RA Donleumann C., Vosberg H.-P.;
 RT "A high risk phenotype of hypertrophic cardiomyopathy associated with

[illegible]

QY 1115 DLFENSVDNECAVSRKCVPRK-SDLGFPPADPSVLVONFNSDFNGKWITSGLN 1173
 Db 1803 HL--NLVLDQNEYSTLREPLKRNVEDLINTLSKLANIAIPNTIYQDLIGKYE-----N 1855
 QY 1174 PTFDAFDCQLHEFHTEGDKLVGNISWRIKTLDSGFTTSRAVQKVFVODPNQPGVLY---- 1229
 Db 1856 PNFDYKLSLAK-----MDYVAISRQDYELMVAKEYKQDYLYKIS 1896
 QY 1230 -----NHDNEYLHYDDWYILSSKIENKPEDYIFVYGRNDADWG 1270
 Db 1897 SEKIDHIVVPLSEYNLMVTNRNPSLSYLKKEKAVLNHILKEDDYKNIL-----AVSE 1950
 QY 1271 YGVAVVYTRSSVPLNSIIPLELEKAASIGRDFSTFIRTDNCTGCEPALVERIEKTVEEGE 1330
 Db 1951 HPTVHLSEHLLKLVLD-----RNDPATMSRSIE-----KPTIDFLSTKALSMG- 1997
 QY 1331 RIIVKEV-----EETEERVEKEVKVGRTEMTLFLQRLAEGFNELKQDBENFVRELSKEEM 1385
 Db 1998 KILVNESTHKNKLLSPDSEFLTMKAKEQGLIISSEKEYSELRLQDIDRPNLDVLEKA 2057
 QY 1386 EFLDIKMEASEVKLFKALP 1407
 Db 2058 AIFDSIIVENIYQQLVNTTSP 2079

RESULT 7
 MYSB_RAT STANDARD; PRT; 1935 AA.
 AC P02564;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
 GN MYH7.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID-10116;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=90016823; PubMed=2798112;
 RA Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
 RT "Complete nucleotide sequence of full length cDNA for rat beta
 RT cardiac myosin heavy chain.";
 RL Nucleic Acids Res. 17:7529-7530(1989).
 RN [2]
 RN DISCUSSION OF SEQUENCE.
 RX MEDLINE=90133919; PubMed=2614840;
 RA McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;
 RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences.
 RT Comparisons suggest a molecular basis for functional differences.";
 RL J. Mol. Biol. 210:665-671(1989).
 RN [3]
 RN SEQUENCE OF 1524-1935 FROM N.A.
 RX MEDLINE=82220036; PubMed=7045682;
 RA Mahdavi V., Periasamy M., Nadal-Ginard B.;
 RT "Molecular characterization of two myosin heavy chain genes expressed
 RT in the adult heart.";
 RL Nature 297:659-664(1982).
 RN [4]
 RN SEQUENCE OF 1871-1935 FROM N.A.
 RC STRAIN=Wistar; TISSUE=Heart;
 RX MEDLINE=85179510; PubMed=6241892;
 RA Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;
 RT "Cardiac myosin heavy chain isozymic transitions during development
 RT and under pathological conditions are regulated at the level of mRNA
 RT availability.";
 RL Eur. Heart J. 5:181-191(1984).
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) 485

CC -1- AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMN) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
 CC MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
 CC -1- MISCELLANEOUS: THERE ARE 10 OR MORE MYOSIN HEAVY CHAIN GENES IN
 CC THE RAT, TWO OF WHICH ARE SPECIFIC FOR ADULT CARDIAC MYOSIN HEAVY
 CC CHAINS.
 CC -1- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
 CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE
 CC CONSERVED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X15939; CAA34065.1; -;
 CC EMBL: J00752; AAA41654.1; -;
 CC EMBL: X32698; AAA41659.1; -;
 CC PIR: S06006; S06006.
 CC PIR: A02989; A02989.
 CC HSP: P08799; IMND.
 CC InterPro: IPR000048; -;
 CC InterPro: IPR001609; -;
 CC InterPro: IPR002928; -;
 CC Pfam: PF00612; IQ; 1.
 CC Pfam: PF01576; Myosin_tail; 1.
 CC Pfam: PF00063; myosin_head; 1.
 CC PRINTS: PR00193; MYOSINHEAVY.
 CC PROSITE: PS50096; IQ; 1.
 CC KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family.
 FT DOMAIN 1 839
 FT GLOBULAR HEAD (S1).
 FT DOMAIN 840 1935
 FT RODLIKE TAIL (S2 AND LMM DOMAINS).
 FT DOMAIN 840 802
 FT COILED COIL (POTENTIAL).
 FT NP_BIND 178 185
 FT ATP.
 FT DOMAIN 655 677
 FT ACTIN-BINDING.
 FT DOMAIN 757 771
 FT METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 129 129
 FT ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 705 705
 FT ALKYLATION (SH-2) (POTENTIAL).
 FT CONFLICT 1529 1531
 FT IRK -> VRR (IN REF. 3).
 FT CONFLICT 1731 1731
 FT D -> H (IN REF. 3).
 FT CONFLICT 1784 1784
 FT N -> K (IN REF. 3).
 FT CONFLICT 1851 1851
 FT T -> N (IN REF. 3).
 FT CONFLICT 1858 1858
 FT R -> K (IN REF. 3).
 SQ SEQUENCE 1935 AA; 223082 MW; C8376C324A7BD82B CRC64;
 Query Match 1.9%; Score 142.5; DB 1; Length 1935;
 Best Local Similarity 16.7%; Pred. No. 2.8;
 Matches 247; Conservative 218; Mismatches 558; Indels 459; Gaps 56;
 QY 166 PDTEQICKGDLFENSVDQNECAVSRKCVPRKSDVGEFVPRDN-----AVV 216
 Db 375 PDGTE-----EADSAVLMGLNSADLLKGLCHPRKVGNEVYTKGVNQVQVAYIGA 428
 QY 217 QNFNMKDFSGKWITSGLNPT-----FDAFDQQLHE---FHMENDKL 255
 Db 429 KSVYKMFN--WNVTR-INATLETKPROYFVIGVLDIAGFEIFDFNSFQLCINFTNEKL 485

SEQUENCE 1025 AA; 116676 MW; FB80378727ED71D8 CRC64;

Query Match 1.9%; Score 139.5; DB 1; Length 1025;
Best Local Similarity 20.9%; Pred. No. 1.8;
Matches 159; Conservative 94; Mismatches 225; Indels 283; Gaps 41;

QY 345 PESIIPNLOKAASVGRDNFFITDSCGPEPPLVERLEKTAEGEKKLLIKEAVEI--- 401
DB PEQENLRLGLVNGKIG-DIDSKVSKASY-----LLUKLEQAHNMKSIVDAIVDIALR 459
QY 402 -----BEEVEKEVKRDTEMTLQRLLEGFKFLQODEEN-- 436
DB 460 PNADYHTTYSVITLQNLKRSVANKLVK---YVFTLFEKFL-----INTDKNTN 511
QY 437 -FVRELSK--EKEILN-----ELQMEATEVE-----KLF-----GRALPIR 470
DB 512 GYVKSNSKSYEEKRKNFKKGKGGKSVKIEKTENEVLDEKNSKLFSGALLTGINRAPPFA 571
QY 471 KLKMA-----LAPHSNF-----LANHETIKYVVGSKLPGHFRFSGWEDYF 511
DB 572 QIPASVYEVHMETLTKITHSNFNFTNTSQALVLIQVTVK-----AKLNSDRYRTLSLF 627
QY 512 GSIVVAKICSSRRIPRYPRKSPRICGDLSDRGLOLFSHGKHNLSPAHSINQVPGNSGC 571
DB 628 D-----PRLVNSS-----KQGIYL-----NLLY 645
QY 572 KPPKDVALLWVEKQGFAPKTAIVAIFILSVASKADAVDALCTCTCLKKECKLELAKCISN 631
DB 646 KSLKQDALNV-ERVEAFYKR-----ILQVCSHLNVGTITGFFFL-----IQLAKTV-- 692
QY 632 PACAAANVACLOTNNRPDETECIKCGDLFENSVDNECAVSRKKCVPKSDVGDPPV 691
DB 693 -----POIKNLIT--NTPVDYE-----YESDAEEGQDKIRKEYDGRKRD----- 732
QY 692 PDPSVLVQKDFMDKFGSKWFTIRGLNFTDADFQCLHEFH-----TEENKLVGNLSWRIRT 747
DB 733 -----PKFAEAKSSULWEINNFN-----HFHPTVKTIANAYVTGETEQIAK 774
QY 748 PDGEGFT-----RSVQV-----KFVODPKYPGILYN-----HDNEYLL 780
DB 775 PDGLGLTLTSLHDFRVYRSKAKOTNARTGTSIMQPLFSGSRVNDVSVLVKASDIMHD-QGPV 833
QY 781 YQDDWYILSSKVEN-SPEDYIFVY-----KGR-----NDAMDGY 814
DB 834 NTEW--LTKKVEDIKPKDEKFFQYFTTKTADGKGNKSNKASNFDSDDEMEINW--- 888
QY 815 GGSVLYTRSAVLPESIIPELQTAQKVGRDPNFTFIKTDNTCGPPLP--VERLEKKVEEG 872
DB 889 -SALVKSRPDVEDSDSSELDFAED-----DP-----SDSTSDEPKLDAIDDEDAKSEGS 938
QY 873 ERTIIEVEEIEEVEKVRDKEVTLSKLFEGFKELQORDENFRELKSEMOVLGLKM 932
DB 939 QES---DQEEGLDEIDFYFDGEQDNSDKK-RSFAESSEDEESSEKEEKEENKEVSAKRA 995
QY 933 EATEVEKFLGRALPIRKLMVATHCFTSPCHDRIRFFSSDD 973
DB 996 KKKQKKNML-KSLPV-----FASADD 1015

RESULT 9

MLP1_YEAST STANDARD; PRT; 1875 AA.
AC 002455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MYOSIN-LIKE PROTEIN MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

SEQUENCE FROM N.A.
STRAIN=S288C;
MEDLINE=93247549; PubMed=8483450;
Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369(1993).
[2]
SEQUENCE FROM N.A.
MEDLINE=94205265; PubMed=8154186;
RX Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UB12 and MPL1 genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L01992; AAA34783.1; -
DR EMBL; X73541; CAA51948.1; -
DR EMBL; Z28320; CAA82174.1; -
DR PIR; S38173; S38173.
DR SGD; S0001803; MLP1.
KW Myosin; Heptad repeat pattern; Coiled coil; DNA repair.
FT DOMAIN 69 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
FT CONFLICT 301 301 R -> A (IN REF.1).
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C9066867 CRC64;

Query Match 1.8%; Score 137.5; DB 1; Length 1875;
Best Local Similarity 17.5%; Pred. No. 5.1;
Matches 252; Conservative 240; Mismatches 525; Indels 421; Gaps 67;

QY 171 CQIKCGDLFNSVVD-----QFNEC-----AVSRKKCVPRKSDVGEPP 208
DB 30 CSLEQVKSFQGVVVKHLNDKLLQFNEKSENKVTVSFDELKASSLKIDGLKTEM-ENV 88
QY 209 VPDENAV-----VQNFNMKDFSGKWIYITSLNP-----TFDAF 241
DB 89 IRENDKIRKERNDFVKFESVENEKMKLSLEFVFRKLDLDTTEKKETQSNQORTLKL 148
QY 242 DCOLHEFHM-----END-----KLVGNTLWRIKTLDCGFFT-----RSVQTFVQDDPLPG 287
DB 149 DERLKEIELVRVNNNSNCKLRTIMDLTKQGGYITNDLNSFTELEKRTQELTL-- 206
QY 288 ALYNHNEFLHYODDWYILSSQIENKPDYIFVYVYGRNDAMDGYGGSVIYTRSPPLPS 347
DB 207 -----LQSNNDW--LEKELRSKNEQVL--SYRQKTD-----K 234
QY 348 IIPNLQKAASVGRDF-----NNFITDNSCGPEPPLVERLEKTAEGEKKLLIKEAVEIE 402
DB 235 VILDIRNELNRLNDFOMERTNNDVLKQKN-----NELSKSIQ--EKLL--EIKGLS 282
QY 403 EEEVEKEVKVRDTEMTLFOR---LLEGFKELQODEENFVREL-----SKEEK 446
DB 283 DSLNSEKQEF-SAEMLKQRLVDLLESQLNVAKEELNSRELNTAKVIADDSKKQTPE 341
QY 447 EILNELQMEATEVEKFLGRALPIRKLR-MALAPHSNFLANHETIKYVVGSKLPGHFRFSW 505
DB 342 DLLKELQLTKEKLAQCEKECEKLRSSITDEADENELSAKSSDDFFLKKQLIKERTKE 401

Qy 506 GWEDYFGSIIVAKISRRIPRYERKSPRCCGLDGRGLQFSGHKNLSPAHSINQNPV 565
Db 402 HLQOIETTFIVLEHKVPIINSKERTDMLNENLNAAALLEHTSNEKNAKVELN--A 458
Qy 566 KGNCGCFKPKDVALMVWEKW-----GQFAKTAIYAIFILSVASKA 605
Db 459 KNOKLVEENDLQTLTKQRLDLCRQIQYLLITNSVNSDGSGLRKEEIQ--FIQIMQED 516
Qy 506 DAV-----DALKTCCTLLKECR--LELAKCISNPACANVACLOTCNNRPDETCQKCG- 658
Db 517 DSTITESDSQKVYTERLVEFKNIQLOE-----KNAELKKVVRNLADKLESKEKSK 568
Qy 659 ---DLFNSVVDENECVAVKRCVPRKSDVD-----FPVPDPSPVLVOKED 702
Db 569 QSLQKIESEVNEAKEAIIPLKS---EKMDLESRIEQLKEELKTSVPNEDASVNT 625
Qy 703 MKDFSGKWTTRGLNPTFDADQOLHEFHTEENKLVNLSWRIRTPDGGFTFRSAQKRV 762
Db 626 IKQ-----LTETKRDLESQVQDLQTRISQIT-----RESTEN-----MSLLNKEI 665
Qy 763 QDPKYPGILYNHONEY-----LLYQDDWYILS-----SKVENSPEYIFVYKYG 806
Db 666 QD-----LYDSKSDISIKLGKSKSRILAEERFKLLSNLTDTKAENDQLRKRFYQLQ 719
Qy 807 -----RNDAMDGYGG-----SVLYTRSAVLPESIPELOTAQKVGGRDFNTFKTD- 852
Db 720 TILKQDSKTHETLNEYVSKSKLSIVETELLNKEE-----QKLRVHLEKNLQKEL 770
Qy 853 NTCQPE-----PPLVERLEKKEVGERTITKEVEIEEVEKVKVDKEVTLFSKL 901
Db 771 NKLSPEKSLRIMVTOLOTKQEREDLEETRSKQCKKIDEDALSSELK-KET-----SOK 826
Qy 902 FEGFKELORDE-----ENFLRELKEEMDVLGDKMEATEVEKLFGRALPRKILMAYA 954
Db 827 DHHKQLEEDNNSNIEWQNKIEALKDYESVITSVDSKQTDIEKIQYKVSLEKEIE-- 884
Qy 955 THCTSPCHDRIFFS-----SDGIGRLGTRKRKRINGTELLKILPPIQASDLRTTG 1006
Db 885 -----EDKIRLHTYVMDETINDDSL-RKELEKSKINLTDAYSOIK--EYKOLYET- 932
Qy 1007 GRSSRPLSAFSGFKGIFDIVPLPSKNELKELTAPLLKLVGLACAFIVPSADAVDA 1066
Db 933 --TSQSLQQTNSKLDESFKDFT-----NOIKNLT-----DEKTS 964
Qy 1067 LKTCACLLKGRICELAKCIANPACANVACLOTCNNRPDETCQIKCGDILFEN-----SV 1121
Db 965 LEDKISLLKEQMENL-----NNELDLQKKMEKEKADFKKISILONNNKEVEAV 1014
Qy 1122 VDFNECAVSKKCVPRKSDLGEPAPDPSPVLVONENISDFNGKWITSGLNPTFFDAFC 1181
Db 1015 KSEY-----ESKLSKTQNDL-----DQOTIYANTAAQNNYEQELQKHADVSKTISELRE 1062
Qy 1182 QLHEFHTEGDKLVGNLSWRIKTLDSGFFTRSAVQKVFQDPNPGVLYNHNDNEYLHYODD 1241
Db 1063 QLHY-----KGQVKTLN-----LSRQLENALKENKS-----WSSQKRESLEQLD 1104
Qy 1242 WYILSSKIENKPE-----DYIFVYVGRNDAMDWDGYG-----AVVYTRSSVLPNSIIP 1289
Db 1105 --LSNSRIEDLSQNKLLYDQIYIYAADKEVNVNSTNGPCLNNILITLRERDILDTKYT 1162
Qy 1290 ELEKAASIGRDESTF-----IRTDNCTGPEPALVERIEKTVTEGERIIVKEVEETE 1341
Db 1163 VAERDAMKLRQKTLMDVQLQDARTKLDNS-----RVEK-----ENHSSIIQHQDDIM 1210
Qy 1342 EEVEKEVEKVGRTMETLQRLAEGFNLKODENFVRELSKEEMFEFLDEIKMEASVE 1399
Db 1211 EKL-NQLNLLRESNITL-----RNEL-ENNNKKELQOSE-----LDKLLQNVAPIE 1255

RESULT 10

DMD_CHICK

ID DMD_CHICK STANDARD; PRT; 3660 AA.

AC P11533;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DYSTROPHIN.
GN DMD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098331; PubMed=3062582;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "Nucleotide sequence of chicken dystrophin cDNA";
RL Nucleic Acids Res. 16:11815-11815(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89210800; PubMed=3072195;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "The chicken dystrophin cDNA: striking conservation of the C-terminal coding and 3' untranslated regions between man and chicken.";
RL EMBO J. 7:4157-4162(1988).
CC -!- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE PLASMA MEMBRANE.
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN, ABP-120, ABP-180, OR BETA-FODRIN).
CC -!- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X13369; CAA31746.1; -.
DR PIR: S02041; S02041.
DR HSP: Q01082; 1AA2.
DR InterPro: IPR000433; -.
DR InterPro: IPR001202; -.
DR InterPro: IPR001589; -.
DR InterPro: IPR001715; -.
DR InterPro: IPR002017; -.
DR InterPro: IPR002349; -.
DR Pfam: PF00307; CH; 2.
DR Pfam: PF00397; WW; 1.
DR Pfam: PF00569; Z2; 1.
DR Pfam: PF00435; spectrin; 22.
DR PRINTS: PR00403; WWDOMAIN.
DR PROSITE: PS00019; ACTININ_1; 1.
DR PROSITE: PS00020; ACTININ_2; 1.
DR PROSITE: PS00021; CH; 2.
DR PROSITE: PS01159; WW_DOMAIN_1; 1.
DR PROSITE: PS00020; WW_DOMAIN_2; 1.
KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
KW Repeat.
FT DOMAIN 1 244 ACTIN-BINDING.
FT DOMAIN 300 3000 26 SPECTRIN-LIKE REPEATS.
FT DOMAIN 3052 3085 WW.
FT DOMAIN 3086 3357 CYS-RICH.
FT VARIANT 1171 1171 MISSING.
FT VARIANT 1869 1869 Q -> H.
FT VARIANT 1885 1885 K -> R.
SQ SEQUENCE 3660 AA; 422874 MW; 85493DAF6D5B6D4A CRC64;

Query Match

1.8%; Score 135; DB 1; Length 3660;

Best Local Similarity 17.6%; Pred. No. 17;

Matches 245; Conservative 196; Mismatches 482; Indels 472; Gaps 64;

QY 71 KDKSOICSDTSPFEIQRDLKRG--TLI-LEKWRQF-----IQIAIVLC 115
 Db 1923 KKEDNANVRAERLSKGAKAVEPTLVQLSKWRDFESKFAFRRLNFAQTQTVLED 1982
 QY 116 TEVIVPRVDALATCACLKECEIELAKCIANPSCAANVACLOTNNRPDETECOIK 175
 Db 1983 TTFVMTESMTVETVPSTYLAEL-LQLLQALSEVEERLNSPVLOAKD-----C 2030
 QY 176 GOLFENSVDQNECAVSKKCVPR-----KSDVGEFPVPDRNAVQNFNMKDF 224
 Db 2031 EDLLKQ-----EELKNIKDCLGRLOGHIDIHSKTPALQSATPRETANIQD-KLTQL 2083
 QY 225 SGKWITSGL--NPTD-----AFDCQLHEPH-----MENDKL-----VGNLT 260
 Db 2084 NSQWEKVNMYRDRQARDKSEKWRLFHCEMKSFNEMWLTETEEKLSRAQTEAGDVGHVK 2143
 QY 261 WR--IKTLDDGGFFTRSVAQTFVQDPLPG-----287
 Db 2144 TKQFLQELQDGI--GROTVVKTLNVNTEGELIEOSSAADANVLKEQLGNLNRWQETCR 2200
 QY 288 -----ALYHNDHFLHYQDWWYILSJOIENKPDYIFVYGRNDAMDGYGGSVIYTR 340
 Db 2201 QLVKRRKRIEENKILSEFOEDLNKLIILWLEETENVIAIPLPGNEDQLRDLCKGVKLVR 2260
 QY 341 SPTLPESIIPLNLOKAASVGRDFNFITDSCGPEPLVERLEKTAEBEGKLLIKAVE 400
 Db 2261 EELLPHKGI--LKRNETGG-----TTLGSASINPERKKHLESTLKEASRRLLKYSRD 2311
 QY 401 IEEVEKEVEKVRDPTMTLFRLLGFEKLEQOODENFVRELSKEKEITNLEOMEATEVE 460
 Db 2312 LPEK-QKEIE-----ILLKDFTELNQO-----INOLTWITPVK 2344
 QY 461 ----KLFR-----ALPIKRLMALAPHSNFLANHETIKYVCGSKLPGHKRFSGWEDYFG 512
 Db 2345 NOLELYNQVGQFARDIKETEA-----VOAKQPNVEE-----2377
 QY 513 STVAKICSSRRIPRYFRKSPRICGLDSRGLQLFSGHKHNLSPAHNSINQWPKNGSGCK 572
 Db 2378 --VLSKGC-----HLYKEK-----ATHPVKKLEDLNADWK 2407
 QY 573 FPKDVALMVEKWKQFATAIIVAFILSVASKADAVDAKLTCTCLLKECLELAKCISNP 632
 Db 2408 AINHILQLKEK-PTFGPALTSQVLT-SQTVAVDT-----QARTKETTSTPT 2456
 QY 633 ACAANVACLOTNNRPDETECOIKCGDLFENSVDDEFNEC-----AVSR--KKCVPRKS 684
 Db 2457 EMPSSV-----LLEVPALADFNKAWAELTDWLSRLDREIKARV 2495
 QY 685 DVGDFPVPDPSVLVQKMDKDFSGKWITRGLNPTFADFQCLHEPTEENKLVGNLSWR 744
 Db 2496 TVGDLDDINDMIKOKANMODLEOR-----RP-----OLDELITAAQ-----NLKNK 2537
 QY 745 IRTPDGFFTRSVAQKF-VQDPKPYGILYNHNDNE-YLLYODDWYILSSKVSNSPEDYIFV 802
 Db 2538 TSNGEARTIITRIEKIOSQDWHVGYLQNRQOLHEMQDKDSTOWLEAKOE---AEQVLE 2594
 QY 803 YKGRNDAMDGYGGSVLYTRSAVLPESI-----IPELQTAQKVGDRFN- 846
 Db 2595 QAKAKLESWK-----EISYTVALKONSELKQFKEITRQWMNIEGVNDVALKPVDRDYS 2650
 QY 847 -----TFIKTONTGCPPEPLVERLEKKEVCERTI-----IKEVE 881
 Db 2651 DTRKVELMTONI-----NATWATINKRVSEREALESALLMLQEFYLDLEKFLAWLTEAE 2706
 QY 882 EIEEVEKVRDKVTL-----FSKLFEFGFKELQD-----PEN---FLRELSKE 922
 Db 2707 TTANVLQDATHKEXTLEDPPQWVRELKMQWQDLQAEIDAHTDIFHNLNENQKILRSLEGS 2766
 QY 923 EMDV-----LDLKMATEVEKLFGRALPKLMVATHCFTSPCHDRIRFFSDDGIGR 977
 Db 2767 EDVILLQRLRLDMMNFRWSELK-----KSLNTRSHLEAST-----DQWKR 2806

QY 978 LGITRKRINGTFLK-----ILPPIQ-----SADLRTTGGSSRPLSAF 1016
 Db 2807 LHLSLOELAWLQKDELKQQAIPGGDIPTVOKQNDVHRTFKRELKTPVIMNALETV 2866
 QY 1017 RSGFS-----KGIDIVPLP--SKNELKELTAPLLKLVLGVLACAF--LIVPSAD---A 1063
 Db 2867 RLFLADOPVGELEKVPPEPDLSPERAQNVTKVLRQADVDVTEWDLNLSADWQKKI 2926
 QY 1064 VDALKTCACLLKGCRIELAKCIANPACAAVACIOTNNRPDETECOIK-----1112
 Db 2927 DDALER-----LQGLQEAEMDELKLRQAEAFKGSWQP 2959
 QY 1113 CGDLFENSVDVDFNECAVSRKKCVPRK-----SDLGEFPAPDPSPVLVQNFNIS---DFN 1163
 Db 2960 VGDLLIDSLODLHLEKVKVYRAEMVPLKEKVHVQVNELAHRAP-PDIQLSPYTLSCLEDLN 3018
 QY 1164 GKWYITSLNPTDAFDCQLHEFHE-GDNKLVGNISWRIKTLDLSDGFFTSAYO---KFV 1219
 Db 3019 TRKWV---LQVADIRTRQJHEAHRDGP-----SQHFLTTSVOGPWERA 3061
 QY 1220 QDPNQPGVLYNHNDNE 1234
 Db 3062 ISPNKVPYYINHETQ 3076

RESULT 11
 MYSF_HUMAN
 ID MYSP_HUMAN STANDARD; PRT; 1937 AA.
 AC PI3535; Q14910;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE MYOSIN HEAVY CHAIN, PERINATAL SKELETAL MUSCLE.
 GN MYH8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=90323631; PubMed=2373371;
 RA Karsch-Mizrachi I., Feghali R., Shows T.B. Jr., Leinwand L.A.;
 RT "Generation of a full-length human perinatal myosin
 RT heavy-chain-encoding cDNA.";
 RL Gene 89:289-294(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=95324556; PubMed=7601129;
 RA Jullian E.H., Kelly A.M., Pompidou A.J., Hoffman R., Schiaffino S.,
 RA Stedman H.H., Rubinstein N.A.;
 RT "Characterization of a human perinatal myosin heavy-chain
 RT transcript.";
 RL Eur. J. Biochem. 230:1001-1006(1995).
 RN [3]
 RP SEQUENCE OF 502-1937 FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=90235862; PubMed=1691980;
 RA Rober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,
 RA Arnold H.H.;
 RT "Identification of three developmentally controlled isoforms of human
 RT myosin heavy chains.";
 RL Eur. J. Biochem. 189:55-65(1990).
 RN [4]
 RP SEQUENCE OF 860-1937 FROM N.A.
 RX MEDLINE=89234168; PubMed=2715179;
 RA Feghali R., Leinwand L.A.;
 RT "Molecular genetic characterization of a developmentally regulated
 RT human perinatal myosin heavy chain.";
 RL J. Cell Biol. 108:1791-1797(1989).
 RN [5]

SEQUENCE OF 1-46 FROM N.A.
 RA Esser K., Tidhar A., Myszkowski M.;
 RT "Isolation and characterization of the human perinatal MHC promoter.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
 CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE, ARE
 CC CONSERVED.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M36769; AAC17185.1; -;
 CC DR EMBL; Z38133; CAA86293.1; -;
 CC DR EMBL; X51592; CAA35941.1; -;
 CC DR EMBL; M35250; ARA36346.1; -;
 CC DR EMBL; AF067143; AAC21557.1; -;
 CC DR PIR; A30220; A30220.
 CC DR HSP; P13538; 2MYS.
 CC DR MIM; 160741; -;
 CC DR InterPro; IPR000048; -;
 CC DR InterPro; IPR001609; -;
 CC DR InterPro; IPR002928; -;
 CC DR Pfam; PF00612; IQ; 1.
 CC DR Pfam; PF01576; Myosin_tail; 1.
 CC DR Pfam; PF00063; myosin_head; 1.
 CC DR PRINTS; PR00193; MYOSINHEAVY.
 CC DR PROSITE; PS00096; IQ; 1.
 CC KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 CC ATP-binding; Methylation; Alkylation; Heptad repeat pattern;
 CC Multigene family.
 CC KW DOMAIN 1 841 GLOBULAR HEAD (S1).
 CC FT DOMAIN 842 1937 RODLIKE TAIL (S2 AND LMW DOMAINS).
 CC FT DOMAIN 842 1937 COILED COIL (POTENTIAL).
 CC FT NP_BIND 181 188 ATP.
 CC FT DOMAIN 658 680 ACTIN-BINDING.
 CC FT DOMAIN 760 774 ACTIN-BINDING.
 CC FT MOD_RES 132 132 METHYLATION (TRI-) (POTENTIAL).
 CC FT MOD_RES 698 698 ALKYLATION (SH-1) (POTENTIAL).
 CC FT MOD_RES 708 708 ALKYLATION (SH-2) (POTENTIAL).
 CC FT CONFLICT 15 15 A -> R (IN REF. 2).
 CC FT CONFLICT 970 970 E -> Q (IN REF. 1 AND 4).
 CC FT CONFLICT 1072 1072 M -> N (IN REF. 3).
 CC FT CONFLICT 1247 1247 N -> H (IN REF. 1 AND 4).
 CC FT CONFLICT 1251 1252 MC -> DGG (IN REF. 3).
 CC FT CONFLICT 1261 1261 E -> G (IN REF. 1 AND 4).
 CC FT CONFLICT 1297 1297 K -> Q (IN REF. 1 AND 4).
 CC FT CONFLICT 1377 1378 KY -> NT (IN REF. 3).
 CC FT CONFLICT 1504 1505 EN -> AH (IN REF. 1 AND 4).
 CC FT CONFLICT 1847 1847 E -> D (IN REF. 1 AND 4).
 CC FT CONFLICT 1914 1914 D -> H (IN REF. 2).
 CC SQ SEQUENCE 1937 AA; 222762 MW; A3BE2D151792E98 CRC64;

Query Match

1.8%; Score 134.5; DB 1; Length 1937;

Best Local Similarity 17.0%; Pred. No. 7.8;
 Matches 280; Conservative 243; Mismatches 597; Indels 529; Gaps 67;
 QY 43 IRSNN-----GYNSFRLETSYKTSFSDSSSHCKDKSQICSDTDFEELQRLDKRGWTLI 98
 DB 239 VRNDSRFRGKFIIRHFGTTGKLASADIETLLLEKSRV-----TFQLK-----281
 QY 99 LEKOWRQIOLA-----IVLVCT-----FVIVPRVD-----AVDALKTC 132
 DB 282 AERSHIYQITSNNKPDLEMLITNPYDYAFVSGEITVPSIDDOEELMATSALDI 341
 QY 133 ACLLEKREILAKCIANPSCAANVACLOTNN-----RPDETCQIKCGDLFNSVVDQFNE 189
 DB 342 LGFTPEEKVSYVLTGAVMHYGNMKFKQKQREQAEPDGTEDVADKA-----AYLOSLNS 395
 QY 190 CAVSRKKCVPRKSDVGEFPVDRNAVQVNFNMKDFSGK-----WYITSLNPT-----237
 DB 396 ADLLKALCYPRVK-VGNEYVTGQGVYNAVAGALAKAVYKMFVLMVTR-INQOLDTK 453
 QY 238 -----PDAFDCQLHE-----FHMENDKLVLGNLTWRIKTLGCGFFTRSAVQ-TF 279
 DB 454 QPRQYFIGVLDIAGFEIFDFNSLEQLCINFTEKLAQGFNHFVLEQEEYKKEGIEWTF 513
 QY 280 VO-DPDL-----PGALYN-----HDNEFLHYQDDWYILSSOENKPPDYI 318
 DB 514 IDFGMDLAACIELIEKPLGIFISILEECMFPRKATDTSFNKLYDQHLGKSNANOKPK---570
 QY 319 FVYVYGRNDWDG-----YGSYVIY-----RSPTLPESIIPNLQKAA-KSVGRDFNNFI 367
 DB 571 -VVGKAEAFHSLIHYAGTVDNITGWLKDNKOPLDNTVVGLYQKSAKMTLASLFSIYA 628
 QY 368 TTDNSCGPEPPLVERLEKTAEGEKLILKEAVEIEEVEKEVEKVDTEMILFORLLBGF 427
 DB 629 SAE-----ADSSAKKGA-----KKGSSFTVSALFRENJLKM 662
 QY 428 KEOQDEENFVREL-----SKEKEILLNLOMEAT-EVEKLFGRALPIR-----470
 DB 663 TNLRSTHPIHVCIIIPNETKTPGMEHELVLHQLRCNGVLGIRICRGFSRILYGDFFK 722
 QY 471 ----KLRLALAPHSNLANHETIKYVVGSKPLGHRKFSGWEDYF---GSIVAKICSSR 523
 DB 723 QRYKVLNASEIPEGQFIDSKKASEKLLASIDIDHTQYKFGHTKVFFKAGLLGLLEMRDE 782
 QY 524 RPRPRFKSPRICCGLSR-----GLOLFSGHKNLSAHSINQNPV 565
 DB 783 KLAQIITRTQAVRGFLMRVEYQKMLQREALFCIQYNVRAFMNVKH--WPWMKLFKKIK 840
 QY 566 KGNSGCKFPKDVAMVWEKMGQFAKT-----AIVAI-----FISVVS 603
 DB 841 PLLKSAETEKEMATKKEE-----FQTKDELAKSEAKRKELEBKMVTLLEKNDLQLOVS 896
 QY 604 KADAV-DALKTCCTLLKEORLELAKCISNPACAANVACLOTNNRPDETCQIKCGDLFE 662
 DB 897 EADSLADAEERCEQLIKN-KIQLE-----AKIKEYTERAEED-----933
 QY 663 NSWDFNECAVSRKK-----CVPRKSDVGEFPVDPSPVLVOKFDMKDFSGKWFITRGLN 717
 DB 934 -----EINAELEAKRKELEDESELEKIDDLLELILAKVEKEK-----971
 QY 718 PTDFAPDCQLHEFHTEENKLVNLSWRIRTPDGGFFTRSAVQKVFQDPKYPGILYNHNE 777
 DB 972 -----HATENK-VKNLTEMAGLDETIAKLSKEKALQE-----THQOTLD 1011
 QY 778 YLLYQDDWILSSKVENSPEDYFVYVYKGRNDWDGYSGLVYTSVAVLPSIPELOTA 837
 DB 1012 DLQAEEDKVNILLTKARTKLEQOV-----DDELSLEQEKRLRM-----DLERA 1054
 QY 838 AQKVGDFNFTFKTONTGPE--PPLVERLEKK-----VB-----870
 DB 1055 KRKLEGLDK--LAQSEDMDMENDKQOLDEKLEKKEFEISNLISKIEDGEQAVEIQKKIK 1112
 QY 871 -----EGERT-----IKEVEFEIEEVEK-----889

Db 1113 ELQARIEELGEEIEAERASRAKAKORSDLSRELEIEISERLEEAGGATSAQVLELNKKREA 1172
QY 890 -----VRDKEVTLFSKLFEGKELQORDEENFLRELSKEEMDVL 927
Db 1173 EFQKLRDLLEATLOHEAVLAARKKHADSMALGEQIDNLRQVKQLEKESKELMET- 1231
QY 928 DGLKMEATEVEKLFGRALPIRKMAVATHCTSPCHDIRIFFSSDDGIGRIGITRKIRING 987
Db 1232 DDLSSNAEISAKAGN-----LEKM-----SLEDQVSEL--KTKBEEQ 1269
QY 988 TELLKILPPIQSADLRITGGSSRLP-----SAPRSFGSKGIFDIPLPFSKNELKELTAPLL 1044
Db 1270 LRL-LOTEAGVSRQLEDEKDALVSQLSRS-----KQASTQOIEELKHOLE 1323
QY 1045 KLVGLVACAFILVPSADAVDAKTCALLKGCRIELAKCTANPACANVACLOTNNRP 1104
Db 1324 EETKAKNALAHALQSSRHDCDLLRQYEEBQGGKAEQALSK--ANSEVAQWRT----- 1376
QY 1105 DETEQIKGDLFENSVDNECAVSRKCKVPRKSDGLGEPFAPDPSVLVQNFNISDFNG 1164
Db 1377 -----KYETDAIQTREELKEAKKLAQRLOEAE-----HVEAVNA 1412
QY 1165 KWTITSGNLPTDAPDCOLHEFTHEGDNKLVGNISWRIKTLDSGFFTRSAVQKFEVQDPNQ 1224
Db 1413 K--CASLEKTKQRLQNEVEDLMLDVE-----RSNAACAALDKQ 1449
QY 1225 PGLVYNHNEVLYHODDYLSSKIKENKPEYIFVYGRNDANDGAGVAVVYTRSSVLP 1284
Db 1450 ----RNFDKVLSEMKQRYEETQAELEASQKE-----SRSLSTELFKVKNVYEEEL- 1495
QY 1285 NSIIEPELEKAASIGRDFSTFRTDNTCGPPEPALVERIEKTVGEGERIIVKEVEIEEVEV 1344
Db 1496 -DQLETARENKNLQOELSDITEQTAEGCKQIHELEKIKKQVEQKEKQIAALEAEASL 1554
QY 1345 EKEVEKVGRTMTLPQ-----RLAEGFNELKODEENFVRELS----- 1381
Db 1555 EHEEGKILRIQLELQVSKSEVDRIKAEKDEEDQLKRHNHRTVETMQSTLDAEIRSRNDA 1614
QY 1382 ----KEEME-FLDEIKWEASEVEKLFQKAL 1406
Db 1615 LRVKKMEGLDMEMEIQLNHANRLAAESL 1643

RESULT 12
MSPL_PLAFK
ID MSPL_PLAFK STANDARD; PRT; 1630 AA.
AC P04932;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P190).
GN MSP-1.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86136024; PubMed=3004972;
RA Mackay M., Goman N., Bone N., Hyde J.E., Scaife J., Certa U.,
RA Stunnenberg H., Bujard H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RA Pan W., Tolle R., Bujard H.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X03371; CAA27070.1; -
CC PIR; A25120; SAZQK1.
CC InterPro; IPR000561; -
CC Pfam; PF00008; EGF 1.
CC Malaria; Merozoite; Polypotein; Repeat; Signal; Glycoprotein;
CC Transmembrane; GPI-anchor.
CC SIGNAL 1 19 POTENTIAL.
CC CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.
CC DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.
CC TRANSMEM 1614 1630 MEMBRANE ANCHOR.
CC CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 755 755 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 774 774 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 835 835 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 911 911 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 955 955 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1049 1049 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1156 1156 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1436 1436 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 1517 1517 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SEQUENCE 1630 AA; 187289 MW; ADBDEC3CE0A46322 CRC64;
Query Match 1.8%; Score 134; DB 1; Length 1630;
Best Local Similarity 18.0%; Pred. No. 6.7;
Matches 246; Conservative 207; Mismatches 473; Indels 444; Gaps 70;
QY 184 VDQNECAVSRKCKVPRKSDVGEFPVDRNAVQNFNMKDFSGKWTITSGNLPTDAPDC 243
Db 240 IENIELIEESKTKIDK-----NKNATKEEKKLYQAYDLS-----IYNK 281
QY 244 QLHEFHENDKLVGNLTWRITLTDGGFFTRSAVQTF--VODP-----DLPGALYNHNE 295
Db 282 QLEEAH-----NLISVLEKRIIDTLKKNENIKELDKINEIKNPPANGNTPTLLDKNK 337
QY 296 FLHYQDDWYILSSQIENKPDY-----IFVYGRNDAMDGYGGSVYITRSPPLPESII 349
Db 338 IEEHEKEIKETAKTFKNIDSLFTDPLEYLYLREKNKID--ISAKVETKESTEPNE-Y 394
QY 350 PNLOKAASVGRDFNPNFTTDSGCGPEPLVERLEKTAEEGKLLI-----KBAVEIEEE 404
Db 395 PNGVTYPLSY-NDINNALNELNSFG-----DLINPFDYTKPSKNITDNERKKFINKKEK 450
QY 405 VEKEVEKVRDTEMTLFORLLEGEKELQODEENFVRELSKEEKEELNEL-----QMEAT 457
Db 451 IKTEKKKI-----ESDKKSYEDRSKSLNDITREYKELKLEIYDSKFNNNIDLT 498
QY 458 EVEKLFGR--ALPIRKURLMALPHSNFLANHETIKYVVGSKLPCHKRFSGWEDY-FGSI 514
Db 499 NFEKMMGKRYSYKVEKL-----THHTTFASYENSKHNL-EKLTALKY---MEDYSLRNI 549
QY 515 VVAKICSSRRIPRYFRKSPRICGLDSRGQLFSGHKHNLSPAHNSINQVPGNSGCG--- 571
Db 550 VWEKEL-----KYKK-----NLISKIENEI---ETLVENIKKEDEQLFEK 586
QY 572 -----KFPKDVLMWWEKQFAKTAIVAIFILSVASKADADYDALKTCCLLKEGRLE- 624
Db 587 KITDKENKPKDEKILEV-----SDIVKQVQVKVLLM-----NKIDELKLTOLILKNVELKH 636
QY 625 -----LAKCISN-----PACAA-----NVACLQTCNNRPDET 651

```

Db 637 NIHVPNSYKQENKQEPYILVLKKEIDKLVFMPKVESLINEEKNKNIKTEGQSDNSEPST 696
QY 652 EQOI-----KCDLFENSVDNECAVSRKCKVPRKSDVDGDFVPDPDSVLVQ-- 699
Db 697 EGEITGOATKPGQAGSALLEG---DSVQAQAQEQQAQBP---PVPVPEAKAQP 747
QY 700 -----KFDMDKDFSGKWFITRGLNPTDFAQCQLHEFTEENKLVGNLSWRI 745
Db 748 TPPAPVNNKTENYSKLDYLE-----KLYEFL-----NTSYIC 779
QY 746 RTPDGGFFTRSAVOKFVQDDPKYGPILYNHD--NEYLLYQDDWILSKVEN-----SPED 798
Db 780 H-----KY--ILVSHSTMENKILQ---YKTKKEESKLSGCDPLD 815
QY 799 YIF-----VYKGRNDAMDGGSV-----LYTRSAVL-----PESIIPELQAAQ 839
Db 816 LFNINQNNIPWY-----SMFDSLNSLSOLFMEIYKEWVCNLYKLKDNKIKNLEAK 871
QY 840 KVGDRFWTFIKTNTCGPEP-PLVERLEKKVBEGRT-----TIKEVEEI----- 883
Db 872 KV-----STSVKTLSSSQMPLSLTPQPKPEVSANDDTSHSTNLSNKLKLFENILSLGKNK 927
QY 884 -----EEVEKVRDKVETLFSKLFEGFKELORDEENFLRSELSKEEMDVLGLK 931
Db 928 NIYQLIGQKSSSENFYEKILKDSFTFNESFTNFVSKADDIINLNDSEKRR-----K 980
QY 932 MEATEVEKFLGRALPIRKLMAVATHCTSPCHDRIRFFSSDDGIGRLGITRKRINGTFLL 991
Db 981 LE-EDINKL-----KKTQLSFDLYNKYKVLKLERLFDKKTGVKVKQIKKLT---LL 1029
QY 992 KILPPIOSADLRTTGGRRSP---LSNFRSGSKGIFDIPLPSKNEKELTAPILLKL 1048
Db 1030 K-----EQLSKLSLNNPKHVQLQNFVFFNK-----KKEAE----- 1061
QY 1049 GVLACAEFLVPSADAVDAKTCALLKGCGR--TELAKCIANPACAAVACLOFCNNRPDE 1106
Db 1062 -----IAETENTLTKILLKHGKLVKYNGESSPLKLTISESIQTDNYASL 1110
QY 1107 TECOIKGDLFENSVDNECAVSRKCKVPRKSDLDGEFPAPDPVSLVQNFNIDSPNGKW 1166
Db 1111 E--NFKVLSKLEGLKLDNLN---LEKKLSYSSGLHLLIA-ELKEVIRKNKY----- 1157
QY 1167 YITSLNPTDFAQCQLHEFTEENKLVGNLSWRIKTLSDSGFFTRSAVQKVFQD----- 1221
Db 1158 ---TGNFSPS-----ENNTDVNNALE---SYK-KFLPEGTDVATVSESGSDLEQS 1201
QY 1222 -PNQPGVLYNHDEYLYHQDDWILSKLENKPEDIFVYVYGRNDA-WDGYGAVVYTR 1279
Db 1202 QPKKPA-----STHVGAESNTITTSQNVDDVDVILVPIFGESEEDYDGLQVY--TG 1253
QY 1280 SSVLPNSIIPLEKAAKSIGRDFSTFRTDTCGPEPALVERTEKTVESGERIIVKEVEE 1339
Db 1254 EAVTPSVI-----DN-----ILSKIE---NEVEVLYLPLAG 1282
QY 1340 IEEVEKEVEKVGRTMTLQRLAEGFNLKQDEENFVRELSKEEMFELD 1389
Db 1283 VYRSLKKOLE---NNVMTFNVNVKDIILNSRFNKNRENFKNVLESDDLIPYKD 1329

RESULT 13
MSPI_PLAPFW STANDARD; PRT; 1639 AA.
AC P04933;
DT 13-AUG-1987 (Rel. 05, Created)
QT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMMSA) (P195).
GN MSP-1.
OS Plasmodium falciparum (isolate Wellcome).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5848;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=86014355; PubMed=2995820;
RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
RA Nicholls S.C., Hillman Y., Davey L.S., Tizard M.L.V., Schwarz R.T.,
RA Freeman R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites.";
RL Nature 317:270-273(1985).
RN [2]
RP REVISIONS.
RA Holder A.A.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X02919; CAA26676.1; -.
CC PIR; A24594; A24594.
CC InterPro: IPR000561; -.
CC Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 20 1639
FT MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 116 116
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 268 268
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 764 764
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 768 768
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 783 783
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 844 844
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 920 920
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 964 964
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1058 1058
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1165 1165
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1174 1174
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1445 1445
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1526 1526
FT N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;

```

```

Query Match 1.88; Score 134; DB 1; Length 1639;
Best Local Similarity 18.08; Pred. No. 6.7;
Matches 246; Conservative 207; Mismatches 473; Indels 444; Gaps 70;

QY 184 VQDFNECAVSRKCKVPRKSDVGEFPVDPDRNAVQVFNKMDFSGKWIYITSLGNTPTDAPDC 243
Db 249 IENINELIEESKKTIDK-----NKNATKEEKKLYQAQYDLS-----IYNK 290
QY 244 QLHEPHMENDKLVGNLTWRIKTLDDGGFFTRSAVQTF--VQDP-----DLPGLYNHDNE 295
Db 291 QLEEAH-----NLISVLEKRRIDTLKKNENIKELLDKINEKNPPPPANSNTPTLLDKNKK 346
QY 296 FLHYODDWYILSSQIENKPDY-----IFVYVYGRNDAMDGSGSVIYTRSPITPESII 349
Db 347 IEEHEKEIKEAKTIKFNIDSFLTDPLELYLREKKNID--ISAKVEIKESTEPNE-Y 403
QY 350 PNLQAAKSVGRDNFNFTTTDSCGPEPLVERLEKTAEGEKKLLI-----KEAVEIEEE 404
Db 404 PNGVTYPLSY-NDINNALNELNSFG---DLINPDFYTKPSKNIYTDNERKKFNEIKEK 459
QY 405 VEKEVEKYRDTMTLQRLLEGGFELQODEENFVRELSKEEKEILNEL-----QMEAT 457
Db 460 IKIEKKKI-----ESDKKSYEDRSKSLNDITKEYEKLNLNLYDSKFNNDILDT 507

```


FT MOD_RES 128 128 METHYLATION (TRI-) (POTENTIAL).
 FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
 FT CONFLICT 966 966 D -> E (IN REF. 2).
 FT CONFLICT 978 978 T -> TE (IN REF. 2).
 FT CONFLICT 986 986 E -> Q (IN REF. 2).
 FT CONFLICT 1008 1014 DLQAEED -> ALEARKT (IN REF. 2).
 FT CONFLICT 1057 1057 D -> Y (IN REF. 2).
 FT CONFLICT 1060 1060 L -> V (IN REF. 2).
 FT CONFLICT 1095 1095 D -> N (IN REF. 2).
 FT CONFLICT 1217 1217 E -> D (IN REF. 2).
 FT CONFLICT 1271 1271 D -> N (IN REF. 2).
 FT CONFLICT 1327 1327 T -> A (IN REF. 2).
 FT CONFLICT 1358 1358 C -> R (IN REF. 2).
 FT CONFLICT 1504 1504 L -> V (IN REF. 2).
 FT CONFLICT 1537 1537 M -> L (IN REF. 2).
 FT CONFLICT 1556 1556 N -> K (IN REF. 2).
 SQ SEQUENCE 1934 AA; 222928 MW; FDBAC58310B057D CRC64;

Query Match 1.8%; Score 133.5; DB 1; Length 1934;
 Best Local Similarity 17.0%; Pred. No. 8.9;
 Matches 251; Conservative 211; Mismatches 525; Indels 493; Gaps 63;

QY 188 NECAVSRKCVPRKSDVGEFFVPDRN-----AVQNFNMKDFSGKWITSLNPT- 237
 Db 390 NSADLLKGMCHPRVKGNEYVTKGNOOVQVYATGALAKSVYKMFN--WMVTR-INATL 446
 QY 238 -----FDAFCQLHE---FHMENDKLVLGNLTWRIKTLDGFFGFFRSVAVQ 277
 Db 447 ETQKQPROYFVGLDIAGFEIFDFNSFEQICINFNTNEKLQOFFNHMFVLEQEEYKKEGIE 506
 QY 278 -TFVQ-----DPDLPGLALY-NHDEFLHYODDW 303
 Db 507 WTFIDFGMDLQACIDLIBKPMRIMSILEECMFPRKATDMTFKAKLYDNHLGKSNFQKPR 566
 QY 304 YTLSSQIENKPPDYIFVYGRND-----AWDGYGSGVYITRSPPLPESLIPLNOKAA-KS 358
 Db 567 NVKGGQEAH---FSLVHYAGTVYDYNILGW-----LQKNKDPNETVGLYKQSSLLK 615
 QY 359 VGRDNFNITDNSGPPPLVERLEKTAEGEKLKIAVEI---EEVEKEVEKVRDTE 416
 Db 616 LSNLFANT-----AGADAP-VDRGKGAKKAGSSSF---QTVSVLHRENKLNKMTNLRST- 664
 QY 417 MTLFQRLLEGFKELQDDEENFVRLSKEEKE-----LLNELOMEAT-EVEKLFGR 466
 Db 665 -----HPHFVRCLIPNETKSPGVMDNPLVMHQLRNCVGLGIRICRKG 707
 QY 467 LP-----IRKRLMALPHSNFLANHETIKYVVGSKLPGRKFRFSWGWDYFGSIV 515
 Db 708 FPNRILYGDFFQVRVIRLNPAAIPEGQFIDSRKGAELKLLSSLDIDHNOYKFGHTKVFFKAG 767
 QY 516 VAKICSSRIPRYPKSPRICGLDLSRGLQFSGKHNLSPAHS---INONVPGNSGCK 572
 Db 768 LIGLLEMRDERLSKRIITRI--QAQSGLLSRMFEKLLERDRSLLYIOWNI-RAFMGVK 824
 QY 573 -FP-----KDVALLMVWEKMGQFAKTAIVAIFILSVASKADAVDAKLT 613
 Db 825 NWPWMKLYFKIKPLKSAETEKENATMK-EFGR-----VKDALEK 864
 QY 614 CTCLLKECRLELAKCISNPACAAVACLOTCNNRPDTECOIKGDLFENS-----VDE 668
 Db 865 SEARRKELEKMWVSLLOF-----KNDLQLOVQAEQDNLADEERQDLIKNKIQLAEARVKE 920
 QY 669 FNE-----CAVSRK---KCVPRKSDVGEFFVPDPVPSVLVQKFDKDFSGKWFIT 713
 Db 921 MTERLEDEEMNAELTAKKRKLEDECSFELKRDIDDLTLAKVEKDK-----967
 QY 714 RGLNPTFDAFCQLHEFTEENKVLGNLSWRIRTPDGGFFTRSAVQKVFQDPKYPGILYN 773
 Db 968 -----HATENK-VKNLTETEMAGLD-----ETIAKLTREKKALQEAHQ 1003
 QY 774 HDNEYLLYQDDWY--ILSSKVE-----NSPDYIFVYKGRND-----AWDGYGSGVLYTRS 823

Db 1004 QALDDLOAEEDKVNTLTQSKVKLEQQVDDLEGSLEQEKVKRMDLERAKRKLGGDLXLTQE 1063
 QY 824 AVLPESIPELOTAQAKVG-----RDF-----NTFIKTDNTCGPE-----PPLVERL 865
 Db 1064 SIM-----DLENDKOOLDEKLKKKDFELNALNARIEDQALGSQQLKELQARIEEL 1117
 QY 866 EKKVEGERT-----IIEVEEIEEEVEK-----889
 Db 1118 EEBL-EAERTARAKVEKRLSDLSRELEETSERLEEAGGATSVQIEMNKKREAFQKMRD 1176
 QY 890 -----VRDKEVTLPFKLPEFGKELQORDEENFLRELKSKEEMDVLDGLKMEAT 935
 Db 1177 LEEATLQHEATAAARKKHADSVAELEGEQIDNLRQVKQLEKEKSEFKLELDD---VTS 1232
 QY 936 EVEKLFGRALPIRKLMVATHCFTSCHDIRIFFSDDGIGRLGIYTRKRLINGTFLFLKILP 995
 Db 1233 NMQIITIKAKANLEKM-----CRT--LEDOMNEHRS-----KAEETQSVNDL----- 1272
 QY 996 PIQADRLTTGGRRSRPL---SAFRSGFGKIGFIDIVPLPSKNEKELKELTAPLLLLKLVGVL 1052
 Db 1273 TSQRAKILOTENGELSRLQDEKEALISQLTRG-----KLTYTQOOLEDLKRL----- 1318
 QY 1053 CAPLIVPSADAVALKYTCACLLKAGRIELAKTANPACAAVACLOTCNNRPDETF--CQ 1110
 Db 1319 -----EEEVKAKNTLAHALQSARHD-----CDLLREQYEETEAKAE 1355
 QY 1111 IKC-----GDLFENSVDDEFNCAVSRKCVPRKSDGFFPAPDPSPVLVQNFN 1158
 Db 1356 LQCVLSKANSEVAQWTKYETDAIQTELEEAKKLAORLODAEE-----1401
 QY 1159 ISDFNGKWIITSGLNTFFDAFCQLHEFHTGDKNLVGNLSWRIKTLDGFFTRSAVQKF 1218
 Db 1402 -----AVEAVNAKSSLEKTKHRLQNETEDLMDV-----VERSNAAAA 1439
 QY 1219 VODPNOPGVLYNHNDNLYLVODDWYILSSKIENKPEYDIFVYVGRNDADWDGCGVAVYT 1278
 Db 1440 ALDKQK---RNFDKILAEWKOKYEESSQSELSOKE-----ARSLST 1478
 QY 1279 RSSVLPSNIPELE-----KAAKSIGRDFTFIRTDNCTGPEPALVERIKTVEEGERI 1332
 Db 1479 ELFKLKNAYEESLEHLETFKRENKNLQEEISDLTEQLGSGTKSIIHELEKIRKQLEAKME 1538
 QY 1333 IVKEVEIEEEVEKEVEKVGRTMTLFO-----RLAEGFNLKQDEENFVRELS----- 1381
 Db 1539 LOSALEEAEASLEHEGNNILRAQLEFNQTKAETIERKLAEKDEEMEQAKRNHLRVVDSLQT 1598
 QY 1382 -----KEEME-FLDEIKMEASEVEKLFGR 1405
 Db 1599 SLDAETRSRNEALRVKMKMEGDLNEMEIQLSHANMAEA 1638

RESULT 15
 LEF_BACAN STANDARD; PRT; 809 AA.
 ID LEF_BACAN
 AC P15917;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE LETHAL FACTOR PRECURSOR (EC 3.4.24.-) (LF).
 GN LEF.
 OS Bacillus anthracis.
 OG Plasmid pXOI.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 ON NCBI_TaxID=1392;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-49.
 RX MEDLINE=90034185; PubMed=2509294;
 RA Bragg T.S., Robertson D.L.;
 RT "Nucleotide sequence and analysis of the lethal factor gene (lef)
 from Bacillus anthracis.";
 RT Gene 81:45-54 (1989).

[2]
RN SEQUENCE FROM N.A.
RA Lowe J.;
RL Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP ZINC-BINDING.
RX MEDLINE=95154669; PubMed=7851740;
RA Kochi S.K., Schiavo G., Mock M., Montecucco C.;
RT "Zinc content of the Bacillus anthracis lethal factor.";
RL FEMS Microbiol. Lett. 124:343-348(1994).
CC -!- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
CC AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE
CC DEATH. LF IS THOUGHT TO BE A LETHAL FACTOR THAT, WHEN ASSOCIATED
CC WITH PA, CAUSES DEATH. LF IS NOT TOXIC BY ITSELF. PA IS THOUGHT TO
CC BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS, THEREBY
CC FACILITATING THE INTERNALIZATION OF LF OR EF.
CC -!- SUBUNIT: SECRETED ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT
CC PROTEINS, A PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN
CC EDEMA FACTOR (EF). NONE OF THESE IS TOXIC BY ITSELF.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B.ANTHRACIS EF
CC AND LF.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M34 (ZINC
CC METALLOPROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M29081; AAA79216.1; -;
CC EMBL; M30210; AAA22569.1; -;
CC PIR; JQ0032; JQ0032.
CC MEROPS; M34.001; -;
CC InterPro; IPR000130; -;
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolyase; Metalloprotease; Zinc; Toxin; Signal; Repeat; Plasmid.
KW SIGNAL 1 33
FT CHAIN 34 809 LETHAL FACTOR.
FT DOMAIN 34 293 PA-BINDING REGION (POTENTIAL).
FT DOMAIN 300 420 REPEATS.
FT METAL 719 719 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 720 720 POTENTIAL.
FT METAL 723 723 ZINC (CATALYTIC) (POTENTIAL).
SQ SEQUENCE 809 AA; 93786 MW; 8C16B4D7277310AE CRC64;

Query Match 1.8%; Score 132; DB 1; Length 809;
Best Local Similarity 17.2%; Pred. No. 3.4;
Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

QY 733 ENKLVGNLSWRIRPTDGGFFTRSAVOKFVQDPKYPGLYNHD---NEYLLODDWY--I 787
DB 105 EMYKAIGG---KIYVGDGITHKISLEALSSEDKKKIKDIYKDALHHEHYVYAKEGEPV 161
QY 788 LSSKVENSPEDYIFYVYKGRNDAMDGYGGSVLYTRSAVLPESIPELQTAQKVGKRDPT 847
DB 162 L---VQSSSEDVVENTERKALN-----VYIEGKILSRDILSKINQPYOKFLDLVLT 209
QY 848 FIKTNTCG-----PEPLVERLEKKVEEGERTIK----- 878
DB 210 IKNASDSGDQLLFTNLKHEPTDFSEVFEQNSVEQVEVFAKAFAYIEPOHRDVLQLY 269
QY 879 -----EVEETEEVEKVRKVTLFKLFEGEKE----- 907
DB 270 APEAFNYMDKFNQEINLSLEELKQRMLSRYEKWEKIKQHYQWSDSLSEBGRLLAKL 329
QY 908 ---LQREENFLRELSKEEMVDLGLKMBATEV---EKLFGRLPI-----RKL 950
DB 330 QIPIEPKDDIIHLSQEEKELKRIQIDSSDFLSTEEKEFLKQLQIDIRDSLSSEKEL 389

QY 951 MAVATHCFTSPCHDIRREFSSDDGIGRLGITRKRINGFTLLKILPPIQASD-----LRTTG 1006
DB 390 L-----NRIOVDSSNP-----LSEK--EKEFLKKLKDIOQYDINORLQDTG 429
QY 1007 GRSRPL-----SAFRSGFSKGIFDIVPLPSKNEKELKELTAPLLKLVG 1049
DB 430 GLIDSPSINLDVRKQVKRDIQIDALLHQSIGSTLYNKIYLYENNMNINLTATL----- 483
QY 1050 VLACAFIYPSADAVALKTCACLLKGCRIELAKCIANPACAANVACLQTCNNRPDETEC 1109
DB 484 -----GADLVDS----- 491
QY 1110 QIKCGDLFENSVVDE--FNECAVSRKKCVPRK---SDLGEFPAPDPSPVLQNFNISDFNG 1164
DB 492 -----DNTKINRGIFNEFKKFKYSISSNYMIVDINERPALDNERLKWRIQLSPDTR 543
QY 1165 KWIYTSGLNPTFDADFQCLHEFHTEGDNKLV--GNISWRIKTLDSGFFTRSAVQKVFQDP 1222
DB 544 AGYLENG-----KLILQRNIGLEIKDVQ----- 566
QY 1223 NQPGVLYNHDNEYLHVQDDWYILSSKIKENKPEDYIFVYVYGRNDAMDGYGGAVVYTR--S 1280
DB 567 -----IIKQSEKEYIRI-DAKVVPKSKIDTKIOE---AOLNINQEWNKALGPLKTKLIT 617
QY 1281 SVLPNSIIPLEKAAKSIGRDFSTFIRTDTCGPEPALVERIEKTVEEGE-RIIVKEV-- 1337
DB 618 FNVHNYASNIVESAYLILNEKNNIQSD-----LIKKVTNYLVDGNGRFVFTDITL 669
QY 1338 -----EETEEVEKEVEKVGKRTMTLQRLAEGNELKQDEENFVRE----- 1379
DB 670 PNIAEQYTHQDEIYEQVHSGKGLYVPESRSILLHGPKSGV-ELRNDSEGFIFHEFGHAVDDY 728
QY 1380 -----LSKEME-----FLDEIKMEASEV 1398
DB 729 AGYLLDKNSDLVTNSKKFDIDIFKEGSNL 758

Search completed: November 6, 2001, 05:15:07
Job time: 646 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 05:13:08 ; Search time 119.14 Seconds
(without alignments)
1568.027 Million cell updates/sec

Title: US-09-075-375A-4

Perfect score: 7495

Sequence: 1 MALSLHTVFLCKEALNLXA.....MEASEVEKLGKALPIRKVR 1412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.unclassified:*
- 13: sp.vertebrate:*
- 14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	2557	34.1	478	10	Q40593	Q40593 nicotiana t
2	2508	33.5	473	10	Q40251	Q40251 lactuca sat
3	2430	32.4	462	10	Q39249	Q39249 arabidopsis
4	1604.5	21.4	472	10	Q9SM43	Q9SM43 spinacia ol
5	1300	17.3	244	10	Q9SJD9	Q9SJD9 arabidopsis
6	1219.5	16.3	289	10	Q9FUS5	Q9FUS5 oryza sativ
7	210	2.8	522	10	Q9SUL3	Q9SUL3 arabidopsis
8	194.5	2.6	2269	5	Q26223	Q26223 plasmodium
9	161.5	2.2	2771	5	Q26216	Q26216 plasmodium
10	156	2.1	1365	2	Q49525	Q49525 mycoplasma
11	156	2.1	1933	13	Q90337	Q90337 cyprinus ca
12	155.5	2.1	886	1	Q29230	Q29230 archaeoglob
13	151	2.0	1055	10	Q9S722	Q9S722 arabidopsis
14	151	2.0	1939	5	Q25662	Q25662 plasmodium
15	150.5	2.0	1332	4	Q9HAW4	Q9HAW4 homo sapien
16	148.5	2.0	2473	11	Q9Q284	Q9Q284 mus musculus
17	147.5	2.0	1785	5	Q9U0P0	Q9U0P0 plasmodium
18	146.5	2.0	1387	5	Q9G276	Q9G276 plasmodium
19	146	1.9	1935	4	Q9H1D5	Q9H1D5 homo sapien

20	146	1.9	2748	3	Q03767	Q03767 saccharomyc
21	145.5	1.9	1116	5	Q9VNH4	Q9VNH4 drosophila
22	144.5	1.9	839	5	Q26024	Q26024 plasmodium
23	144	1.9	2166	2	Q51465	Q51465 borrelia bu
24	144	1.9	3899	4	Q9Y6Y2	Q9Y6Y2 homo sapien
25	143.5	1.9	1558	5	Q96275	Q96275 plasmodium
26	143	1.9	1978	5	Q9W0M1	Q9W0M1 drosophila
27	142.5	1.9	880	1	Q9UZC8	Q9UZC8 pyrococcus
28	142.5	1.9	1199	5	Q91349	Q91349 caenorhabdi
29	142.5	1.9	1819	2	Q9ZLV0	Q9ZLV0 helicobacte
30	142.5	1.9	5105	5	Q61201	Q61201 caenorhabdi
31	141.5	1.9	1935	6	Q9GKR1	Q9GKR1 sus scrofa
32	141	1.9	1002	2	Q66583	Q66583 aquifex aeo
33	140.5	1.9	1302	2	Q49547	Q49547 mycoplasma
34	140.5	1.9	1930	13	Q9BGD5	Q9BGD5 pennania ar
35	140	1.9	800	1	Q59066	Q59066 methanococc
36	140	1.9	3595	4	Q9UQH3	Q9UQH3 homo sapien
37	139	1.9	2712	10	Q9SB74	Q9SB74 arabidopsis
38	139	1.9	3911	4	Q99996	Q99996 homo sapien
39	138.5	1.8	1327	4	Q9Y2L2	Q9Y2L2 homo sapien
40	138	1.8	1057	10	Q9FI17	Q9FI17 arabidopsis
41	138	1.8	2867	5	Q9N2M3	Q9N2M3 plasmodium
42	137.5	1.8	3908	4	Q9UQQ4	Q9UQQ4 homo sapien
43	137	1.8	1088	4	Q60772	Q60772 homo sapien
44	137	1.8	1109	6	Q00756	Q00756 oryctolagus
45	137	1.8	1676	10	Q23332	Q23332 arabidopsis

ALIGNMENTS

RESULT 1
Q40593 ID Q40593 PRELIMINARY; PRT; 478 AA.

AC Q40593;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN TVDEL.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;

RP SEQUENCE FROM N.A.
RC STRAIN=XANTHI; TISSUE=LEAF;
RA Bugos R.C., Yamamoto H.Y.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U34817; AAC50031.1;
DR Mendel; 9222; Nicta; vdel:9222.
DR InterPro; IPR000566; -;
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Transit peptide.
FT TRANSIT 1 134 POTENTIAL.
FT CHAIN 135 478 POTENTIAL.
SQ SEQUENCE 478 AA: 54561 MW: 0967DF4547D7809D CRC64;

Query Match 34.1%; Score 2557; DB 10; Length 478;
Best Local Similarity 100.0%; Pred. No. 6,9e-138;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	474	MALAPHSNFIANHETIKYYVGSKLPGHKRFSWGWDYFGSIVVAKICSSRRIPRYFKSP	533
Db	1	MALAPHSNFIANHETIKYYVGSKLPGHKRFSWGWDYFGSIVVAKICSSRRIPRYFKSP	60
QY	534	RICCGLDRLQLFSGHKHNLSPAHSINQNPVKGNSGCKFPKDVLMVWEKQCFAKTAI	593
Db	61	RICCGLDRLQLFSGHKHNLSPAHSINQNPVKGNSGCKFPKDVLMVWEKQCFAKTAI	120
QY	594	VAIFILSVASKADAVDAKLTCTCILLKRELAKEISNPACAAVACLQTCNNRPDETEC	653

Db 121 VAIFILSVASKADAVDAKTCCTCLLKECLELAKCISNPACANVACLQTCNNRPDETEC 180
QY 654 QIKCGDLFENSVDDEFNECAVSRKKCVPRKSDVGDFFVPDPSVLVQKDFMKDMSGKWFIT 713
Db 181 QIKCGDLFENSVDDEFNECAVSRKKCVPRKSDVGDFFVPDPSVLVQKDFMKDMSGKWFIT 240
QY 714 RGLNPTDADFDCOLHEPHTENKLVGNLSWRIRTPDGGFFTRSAVOKFVQDPKYPGLILYN 773
Db 241 RGLNPTDADFDCOLHEPHTENKLVGNLSWRIRTPDGGFFTRSAVOKFVQDPKYPGLILYN 300
QY 774 HDNEYLQYQDDWYILSSKVENSPEDYIFVYVYKGRNDAMDGYGGSVLYTRSAVLPESIPE 833
Db 301 HDNEYLQYQDDWYILSSKVENSPEDYIFVYVYKGRNDAMDGYGGSVLYTRSAVLPESIPE 360
QY 834 LQTAQKVGDRFNTFIKTDNCGPEPLVERLEKKVVEGERTIIKEVEEIEEVEKVRDK 893
Db 361 LQTAQKVGDRFNTFIKTDNCGPEPLVERLEKKVVEGERTIIKEVEEIEEVEKVRDK 420
QY 894 EVTLFSKLFEGFKELQDEENFLRELKSEMDVLDGLKMEATEVEKLFGRALPIRKL 950
Db 421 EVTLFSKLFEGFKELQDEENFLRELKSEMDVLDGLKMEATEVEKLFGRALPIRKL 477

RESULT 2
Q40251 ID Q40251 PRELIMINARY; PRT; 473 AA.
AC Q40251;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN VDEL.
OS Lactuca sativa (Garden lettuce).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
OC euasterids II; Asterales; Brassicaceae; Lactuca.
OX NCBI_TaxID=4236;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ROMAINE;
RX MEDLINE=96270536; PubMed=8692813;
RA Bugos R.C., Yamamoto H.Y.;
RT "Molecular cloning of violaxanthin de-epoxidase from romaine lettuce
and expression in Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:6320-6325(1996).
DR EMBL; U31462; AAC49373.1; -;
DR Medel; 8691; Lacs; Vdel; 8691.
DR InterPro; IPR000566; -;
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Transit peptide.
FT TRANSIT 1 125 POTENTIAL.
FT CHAIN 126 473 VIOLAXANTHIN DE-EPOXIDASE.
SQ SEQUENCE 473 AA; 54447 MW; 1B22522DC2C62699 CRC64;

Query Match 33.5%; Score 2508; DB 10; Length 473;
Best Local Similarity 100.0%; Pred. No. 4.2e-135;
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALSLHTVFLCKEALNLYARSPCNERPHRSQOPPTNIIMKIRSNNGYFNFLFTSYK 60
Db 1 MALSLHTVFLCKEALNLYARSPCNERPHRSQOPPTNIIMKIRSNNGYFNFLFTSYK 60
QY 61 TSSFSDSHCKDKSQICSDITSFEEIQRFDLKRGMTLILEKQWQFQLAIVLVCTFVIV 120
Db 61 TSSFSDSHCKDKSQICSDITSFEEIQRFDLKRGMTLILEKQWQFQLAIVLVCTFVIV 120
QY 121 PRVDAYDAKTCACLLKCELELAKCIANPSCANVACLQTCNNRPDETECOIKCGDLFE 180
Db 121 PRVDAYDAKTCACLLKCELELAKCIANPSCANVACLQTCNNRPDETECOIKCGDLFE 180
QY 181 NSVVDQFNECAVSRKKCVPRKSDVGDFFVPDPRNAVQFNMKDFSGKWIITSGLNPTFDA 240

Db 181 NSVVDQFNECAVSRKKCVPRKSDVGDFFVPDPRNAVQFNMKDFSGKWIITSGLNPTFDA 240
QY 241 FDCQLHEFHMDKLVGNLTWRIKTLDGGFFTRSAVQTFVQDPLPGALYNHNEFLHYQ 300
Db 241 FDCQLHEFHMDKLVGNLTWRIKTLDGGFFTRSAVQTFVQDPLPGALYNHNEFLHYQ 300
QY 301 DDMYILSSQIENKPDYIFVYVYKGRNDAMDGYGGSVYVTRSPITLPESIPNLQKAASVG 360
Db 301 DDMYILSSQIENKPDYIFVYVYKGRNDAMDGYGGSVYVTRSPITLPESIPNLQKAASVG 360
QY 361 RDNFNPTDNDSCGPEPLVERLEKTAEGEKKLLIKEAVEIEEVEKVRKOTMTLF 420
Db 361 RDNFNPTDNDSCGPEPLVERLEKTAEGEKKLLIKEAVEIEEVEKVRKOTMTLF 420
QY 421 ORLLGEGFKELQDEENFVRELKSEKEILNLOMEATEVEKLFGRALPIRKL 473
Db 421 ORLLGEGFKELQDEENFVRELKSEKEILNLOMEATEVEKLFGRALPIRKL 473

RESULT 3
Q39249 ID Q39249 PRELIMINARY; PRT; 462 AA.
AC Q39249;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN AVDEL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; euroids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Bugos R.C., Yamamoto H.Y.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U44133; AAC50032.1; -;
DR EMBL; AC003981; AAF99753.1; -;
DR Medel; 6341; Arath; Vdel; 6341.
DR InterPro; IPR000566; -;
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 113 POTENTIAL.
FT CHAIN 114 462 VIOLAXANTHIN DE-EPOXIDASE.
SQ SEQUENCE 462 AA; 52017 MW; 58E37B2C12D4426B CRC64;

Query Match 32.4%; Score 2430; DB 10; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.1e-130;
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 951 MAVATHCFTSPCHDRIRFFSSDDGIGRLGITRKRINGTLLKILPPIQSADLRTTGGRSS 1010
Db 1 MAVATHCFTSPCHDRIRFFSSDDGIGRLGITRKRINGTLLKILPPIQSADLRTTGGRSS 60
QY 1011 RPLSAFRSFGSKGIFDIVPLPSKNELKELTAPLLKLVGLVACAFIVPSADAVDAKTC 1070
Db 61 RPLSAFRSFGSKGIFDIVPLPSKNELKELTAPLLKLVGLVACAFIVPSADAVDAKTC 120
QY 1071 ACLLKGCRIELAKCIANPACANVACLQTCNNRPDETECOIKCGDLFENSVDNECAV 1130

```

121 121 ACLLKGCRIELAKCIANPACAAVACLTQCNRRPDETECOIKCGDLFNSVDFNECAV 180
1131 1131 SRKKCVPRKSDLGEPAPDPSPVSVQNFNSDFNKGWITSGLNPTDFADFCOLHEFHTEG 1190
181 181 SRKKCVPRKSDLGEPAPDPSPVSVQNFNSDFNKGWITSGLNPTDFADFCOLHEFHTEG 240
1191 1191 DNKLVGNISWRITKTLDSGFFTSRAVQKFPVQPNQPGVLYNHDNEYLHYQDDWYILSSKIE 1250
241 241 DNKLVGNISWRITKTLDSGFFTSRAVQKFPVQPNQPGVLYNHDNEYLHYQDDWYILSSKIE 300
1251 1251 NKPEDYIFVYRGRNDWDGCGAVVYTRSSVLPNSIIPELEKAASIGRDFSTFIRTDN 1310
301 301 NKPEDYIFVYRGRNDWDGCGAVVYTRSSVLPNSIIPELEKAASIGRDFSTFIRTDN 360
1311 1311 TCGPEPALVERIEKTEVEGERIIVKEVEIEEVEKEVKGRTMTLFORLAEGFNEKL 1370
361 361 TCGPEPALVERIEKTEVEGERIIVKEVEIEEVEKEVKGRTMTLFORLAEGFNEKL 420
1371 1371 QDBENFVRELSKEEMFDEIKMEASEVEKLFPGKALPIRKVR 1412
421 421 QDBENFVRELSKEEMFDEIKMEASEVEKLFPGKALPIRKVR 462

RESULT 4
Q9SM43 PRELIMINARY; PRT; 472 AA.
AC Q9SM43:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN SVDEL.
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
OC Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Emanuelsson A.K., Eskling M., Akerlund H.E.:
RT "Cloning and sequencing of Spinacia oleracea violaxanthin de-
RT epoxidase."
RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AJ250433; CAB59211.1; -.
KW Signal.
FT SIGNAL 1 124 POTENTIAL.
FT CHAIN 125 472 VIOLAXANTHIN DE-EPOXIDASE.
SQ SEQUENCE 472 AA; 53658 MW; 12B4C8C69AB61E12 CRC64;

Query Match 21.4%; Score 1604.5; DB 10; Length 472;
Best Local Similarity 66.4%; Pred. No. 11e-83;
Matches 299; Conservative 63; Mismatches 69; Indels 19; Gaps 5;

QY 968 FFSSDDGIGRLGTRKINGTFLKILPPIQSADLRTTGRSSRPLSAFSS-----GFSK 1022
DB 37 FLFQDARRNIRFNRDLKSLCTKFGASEKLQ-----HSKSPKSGLISCGWEVNSSK 87

QY 1023 GIFDIVPLPSKNEKELTAPLLKLVGLVACAFIIVPSADAVALKTCACLLKGCRIELA 1082
DB 88 VVSNAV-IPKKWNLKTKL---VVEVTAIVACTFVMSAQAVALKTCCTCLLKECRIELA 143

QY 1083 KCIANPACAAVACLTQCNRRPDETECOIKCGDLFNSVDFNECAVSRKKCVPRKSD 1142
DB 144 KCIANPACAAVACLTQCNRRPDETECOIKCGDLFANKVDFNECAVSRKKCVPRKSDV 203

QY 1143 GEFPAPDPVSVQNFNSDFNKGWITSGLNPTDFADFCOLHEFHTEGDKNLVGNISWRI 1202
DB 204 GEFPAPDPVSVQNFNSDFNKGWITSGLNPTDFADFCOLHEFHTEGDKNLVGNISWRI 262

QY 1203 KTLDSGFFTSRAVQKFPVQPNQPGVLYNHDNEYLHYQDDWYILSSKIEKPEDYIFVYR 1262

```

```

263 263 KTPDGGFFTSRAVQKFPVQPNQPGVLYNHDNEYLHYQDDWYILSSKIEKPEDYIFVYR 322
1263 1263 GRNDWDGCGAVVYTRSSVLPNSIIPELEKAASIGRDFSTFIRTDNCGPEPALVERI 1322
323 323 GRNDWDGCGAVVYTRSSVLPNSIIPELEKAASIGRDFSTFIRTDNCGPEPALVERI 382
1323 1323 EKTVEGERIIVKEVEIEEVEKEVKGRTMTLFORLAEGFNEKLDEENFVRELSK 1382
383 383 EKTVEGERIIVKEVEIEEVEKEVKGRTMTLFORLAEGFNEKLDEENFVRELSK 442
1383 1383 EMEFDEIKMEASEVEKLFPGKALPIRKVR 1412
443 443 EMEFDEIKMEASEVEKLFPGKALPIRKVR 472

RESULT 5
Q9SJD9 PRELIMINARY; PRT; 244 AA.
AC Q9SJD9:
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE T27G7.23 (FRAGMENT).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T27G7 from chromosome
RT I."
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AC006932; AAF22898.1; -.
DR InterPro: IPR000566; -.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
FT NONTER 244 244
SQ SEQUENCE 244 AA; 26631 MW; 4CD8221F444CF04C CRC64;

Query Match 17.3%; Score 1300; DB 10; Length 244;
Best Local Similarity 100.0%; Pred. No. 9.8e-67;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 951 MAVATHCTSPCHDRIRFFSSDDGIGRLGTRKINGTFLKILPPIQSADLRTTGRSS 1010
DB 1 MAVATHCTSPCHDRIRFFSSDDGIGRLGTRKINGTFLKILPPIQSADLRTTGRSS 60

QY 1011 RPLSAFRSGFGKIFDIVPLPSKNEKELTAPLLKLVGLVACAFIIVPSADAVALKTC 1070
DB 61 RPLSAFRSGFGKIFDIVPLPSKNEKELTAPLLKLVGLVACAFIIVPSADAVALKTC 120

QY 1071 ACLLKGCRIELAKCIANPACAAVACLTQCNRRPDETECOIKCGDLFNSVDFNECAV 1130
DB 121 ACLLKGCRIELAKCIANPACAAVACLTQCNRRPDETECOIKCGDLFNSVDFNECAV 180

QY 1131 SRKKCVPRKSDLGEPAPDPSPVSVQNFNSDFNKGWITSGLNPTDFADFCOLHEFHTEG 1190
DB 181 SRKKCVPRKSDLGEPAPDPSPVSVQNFNSDFNKGWITSGLNPTDFADFCOLHEFHTEG 240

QY 1191 DNKL 1194
DB 241 DNKL 244

```


Query Match 2.6%; Score 194.5; DB 5; Length 2269;
Best Local Similarity 18.4%; Pred. No. 0.017;
Matches 255; Conservative 213; Mismatches 436; Indels 483; Gaps 65;

QY 304 YILSOIENKPDYIFVYGRNDADWDGYSVYIYTRS-----PTLPEIIPNLQ--KA 355
DB 34 YISNQIKNKLN--VSTYPEGR---EGTSSLELAKSWEKTKLETITELTKSNEVRL 87
QY 356 AKSVGRDENNITDNSCGPEPPVERLEKTAEGEKL-----LIKEAVEI 401
DB 88 EKEIRELEFKY-----LDEAEKYLEGLLELKKLIDIAKIEYVKNVEL 135
QY 402 EEEVEK-----EVEVRDTEMFLFORLEGFKELQ--DEENFVRELSKEEK 446
DB 136 KKEIEKNAYIDELANQSPYKVTGYENKNTIYNTIKSYFDQIYEGDIDTFYNELSVIK 195
QY 447 E-ILNLOMEATEVEKLEGRALPI--RKLRLMALAPHNFIANHET-----IKYVV 493
DB 196 EDPIIDIE-DKTKLENLASKIDNYDKIQKMEIETVKSHLNIIETNNKLPNTILEIKKI 254
QY 494 GSKLPGHRFSWGEDYFGSIWAKICSSRIPIRYFRKSPRICGLDSRGLQFSPH--- 549
DB 255 YDEI--SKELNMLEDFNK-----EKLSNKISDYDKKREQL-SEYKSKMLEIRNHNSQ 307
QY 550 -----GKHNLSPAHSINQNV-----KNSGCKFPKDV 577
DB 308 TNVDNTEKEAKQNYDKSNEHMTIPTNEDEISKIISVETKMDKDEILSKVNTYIDFNKY 367
QY 578 ALMWKKGQFAPKTAIFAIFILSVASKADAVDAKTCCLLKECRLELAKISNPACAA 637
DB 368 KETVNSEHSGQTE-----LTDKIKAEVSDK-----ELK----- 396
QY 638 VACLQTCNNRP---DETECOIKCGDLFEN---SVVDFNECAVSRKCKVPRKSD----- 685
DB 397 --CEGSPDNKSLNETNSIE--KEYQNIINTLKKVDEYIKVCKSTRESITKFSKOTIL 452
QY 686 -----VGFPPVPPSVLVOKFDMKDFSGKWFITRGLNPTFDAPDCOLHEFHTEE- 734
DB 453 KDLNQNIKTVKETSIDKS-YIEKFE-QILTGR--QTKLENKFTFELSNHEANNEL 507
QY 735 ---NKLGVNLSWRIRTPDGGFFTSRAVQKFPQKPGILYNHNDVLLYQDDWYIUS 790
DB 508 IKYFSDKANLINEENMLYNQFTEK--EKTFDNIKEKNI---HINEEI-----S 552
QY 791 KVENSPEDYIF-----VYKGRNDAMD-----GY 814
DB 553 KIEIKIHASIYINISBETEREIGINIESLNTKVFKEKVENVTNLNKKIKELKHDFSDFGK 612
QY 815 GGSVLYTRS-----AV-----LPEIPIELOTAQKVGDRDNFTFI 849
DB 613 EGNIKYTDKIKKINDIMAVSQIDQHINGLDDIQKKSSEYVSEMKEQINKLEKVSNT 672
QY 850 KTDNTCG--PEPPLVERLEKK--VEEGERTIIEVEIEEE---VEKVRDKEVTLPSKL 901
DB 673 SNDNVEGKKQQQIIVTKIDKKNIYEINKLLSEIKEDNTSLEKVDINLSYGNL 732
QY 902 FEGFELQDENFRELKSEEM-----DVLGLKMEATEVEKLFGRALPIRKLMAV--AT 955
DB 733 GNLFILE-QIDEKKAENTIKSMEAY IDDLNKKKSQEIETEMDIRKMDINKMEALKIS 791
QY 956 HCTFSPCHDIRFFSDGIGRLGTRKTRNGTFLKLLPIQISADLRTTGGRSSRPLSA 1015
DB 792 HDDDKKCHDKSK--NHKENISDIYDKSKIIQDF-----SRESNDINKNLQKNVSE 842
QY 1016 FRSGFS-----KGIFDIVLPFSKNELKEL-----TAP 1042
DB 843 SQNHNSDINQCANEVANYINILK---NKIRKIDKVEYKTESEKKNKNTINDELNSEK 899
QY 1043 LLLKLVGLVACAFIVPSADAVDAKTCACLLKGCRIELAKCIAN---PACAAVACLQ 1099
DB 900 VIKKTEGLS-----LKECRSINSTLDDKDIIDEIKINVLKK 938

QY 1100 CNRPDETEC--QIKCGDLFENSVDDEFNEC----- 1128
DB 939 -NILNEETNITNHNFAEEYVKIVLSNFNNIEMADNKSQVILEIKKNGTNDHDYNIKEL 997
QY 1129 -----AVSRKKCVPRKSDLGFPAPDPSPVLQVONENISDFNGKWTITSLNPT 1175
DB 998 KSHKDSNGYKTEADONKRAIOKNKELFYQKEEVTVLLNKYYAVELKKN----- 1047
QY 1176 FADFQDLQHEFHTEGDKLVGNISWRIK-----TLDSGFFTSRAVQKFPQKPGVLY 1229
DB 1048 ---FD-----KTKNDK--QIIKEIKADHNYCTLESG-----KSEKKMNE----- 1082
QY 1230 NHDNEYLHQQDDWYILSSKIENKPEYIFVYVGRNDAMDWDGAGVAVYTRSSVLP----- 1284
DB 1083 -IKNEKIHIHDE--VANNDKSNK-----AITSIKSVPEPKTKI 1118
QY 1285 ---NST-----IPELEKAAKSIGRDFSTFIRTDNCTGPEPALVERIEKTVVEGR 1331
DB 1119 IKNEIRTSDDCLKTEINDLEKIOISLSDTQETKUTEN--GKQLTLEELLESKKQK- 1175
QY 1332 IIVKEVEIEEVEEKEVKGRTMTLQF-----LAEGFNEKQDEENFY---RELS 1381
DB 1176 ---KNIEDQKKELDEVNSKIKNIENVTNQHKKNYEGIVEKINEIAKTNNKQIESTKELI 1232
QY 1382 KEEMFEL 1388
DB 1233 KPTIOHI 1239

RESULT 9
Q26216 PRELIMINARY; PRT; 2771 AA.
AC Q26216;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE RHOPTRY PROTEIN.
OS Plasmodium yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5861;
RN [1]
RP SEQUENCE OF 379-2771 FROM N.A.
RC STRAIN=YM;
RX MEDLINE=97077455; PubMed=8920022;
RA Sinha K.A., Keen J.K., Ogun S.A., Holder A.A.;
RT "Comparison of two members of a multigene family coding for high-
RT molecular mass rhoptry proteins of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 76:329-332(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=YM;
RT "Structure of the E8 gene encoding a high molecular mass rhoptry
RT protein of Plasmodium yoelii.";
RL Mol. Biochem. Parasitol. 0:0-0(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=YM;
RA Holder A.A.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U36927; AAB41263.3; --
SQ SEQUENCE 2771 AA; 325640 MW; C0CCB9AB6E7ACF36 CRC64;

Query Match 2.2%; Score 161.5; DB 5; Length 2771;
Best Local Similarity 16.5%; Pred. No. 1.7;
Matches 250; Conservative 269; Mismatches 481; Indels 519; Gaps 68;

QY 59 YKTSFSDSHCKDKSQICSIDTSFEEIQRLKRGMTLILEKQWQFTQLAIVLVCTFV 118
DB 857 YKEVESEHNKFTLTNKTKEVSDDEIKYENKFNDSKSLINETKKSIETEEYQNT-- 914

```
Qy 119 IPRVDAVDAKLTACLLKECRIELAKCIANPSAANVACLOTNNR-----PDTECOIK 174
Db 915 -LKKVD--DYIKVC-----LNTNELITN-----CHNQTTLKDKLNQNIK 951
Qy 175 C-----GDLFNSVVDQ-----FNECAVSRKKCV-----PRKSDVGEF 207
Db 952 TIKETNSIDKIYTKDFENILTDKKTELETFTGLSLNNHESNNKELTYFYDLKANLKG- 1010
Qy 208 PVPORNAVQVFNKDFSGKWYITSGLNPTDFADFCOLHEFHENDKLGNLTWRKTLTD 267
Db 1011 --NKENMLYKQFNEKEAVE-----DIKKKNVDINKIVSNIETIYT-- 1050
Qy 268 GGFTSRSAVOTFVODPLPGALYNHDN-----EFLHYODDWYLS-----S 308
Db 1051 -----SIYNINEDTENEIGKSTELLNTK-----VLEKVKANVTNLN 1086
Qy 309 QIENKPPDYIFVYVYGRNDWDGCGSVYITRSTPLPESIIIPNLOKAASVGRDFNNFIT 368
Db 1087 EIKELKDYDFQDP-----GKEKNIKYPDENKIKNDIDTLNQRKDKSIETLTETKKN 1138
Qy 369 TDNSCGPEPPLVERLEKTAEGEKLLIKEAVEIEEVEEVEKEVRODEMFLQO---RLLE 425
Db 1139 SENHIDEIKGIDKLKVV--PNKTFWEDPKEIEKKIENIVEKI-DKKKIYKEIDKLLN 1195
Qy 426 GFKELOQDEEN-----FVRELSKEKEI-----LNELOM 454
Db 1196 EISKIENDKTSLEKLNINLSYKSLGNLFLOQIDEEKKAETHIKAMEAYIDLOLNKIK 1255
Qy 455 EATEVEKLFGRALPI-----RKLMLAPLAPHSNLANHETIKIYVVGSKLPGHKRFS 504
Db 1256 KSOIEKEMINMDIKMDIHEKMKALNISDDDYKIYHTTSKNH--EEKISDIRKNSKIIQ 1314
Qy 505 -WGWDYFGST-----VVAKISGRRIPRYFRKSPRICGLDSRGLQFSGHKHNLSP 556
Db 1315 DFSESYINDIKKEKNVLESQNNNTDINOYLSKIENI-----YNILK 1358
Qy 557 AHSINQNVPGNSCKPKPDVALMVWFKQGFQAKTAIVAIFILSVASKADAVDAKLTCTC 616
Db 1359 LNKTKKIIDK-----VKYTDIEKN-----NKKINAELSSEKIIITQ 1396
Qy 617 L-----LKECRLELAKISN---PACAAVACIQC-----NNRPDETCQIKCGDLFEN 663
Db 1397 LKENSLSKEQSKIKSTIDDNYVSECINKNITLKIYIVNEKN-----INIFYKN 1446
Qy 664 SVVDFNECAVSRKKCVPRKSDVGVDPVDPDSVLVQKDFMKDPSGKWFITRGLNPTFD 723
Db 1447 A--EYINQ-----NVSLFNFIEMADTKSQYIL-----1472
Qy 724 DCQLHEFTEENKLVGNLSWRIRTPDGGFTFSVAVQKVFQDPKYPGILYNHDNEYLLYQD 783
Db 1473 -----NKKNGTNTDYNKE-----LKEHKKSNNYIKD 1502
Qy 784 DWYILSKSVNSPEDYIFVYVYGRNDWDGCGSVLYTRSAVILPESIIPELOTAQKVR 843
Db 1503 E-----AGNTOE-----IKKNKELFEYQEV-----TVLLNKYVAVELK----- 1538
Qy 844 DFNFTIKDTCGPEPPLVERLEKKEVEGERTIIEVEEIEEVEEVEKVRDEVTLSKLFEE 903
Db 1539 --NKFDRTKN-----YSQIIEKIDKAHTFTTSQADSKEMKNEIKESQIRI----- 1583
Qy 904 GFKELOQDEENFLRELSKEMDVID-GLKMEATEVEKLFGRALPIRKLMAVATHCF--TS 960
Db 1584 -----EDEAVKNNKSNKAILDITLSVEPKI-----KFLKIDKLTKSDCLKETK 1629
Qy 961 PCHDRIRFFSDGIGRLGTRKRNKINGTFLLLKILPPIQSADLTGTGGRSSRPLSAFRSG 1020
Db 1630 DIETKINLSIDTQETKL-TEKNILNT-LEKLLLESKNO-----1667
Qy 1021 SKGIFDIVPLPSKNEKELTAPL-----LLKLVGLACAFIIVPSADA 1063
Db 1668 KKNTE-----QKDELDEVSKIKNTESNVQHKNYEIGIVEIKINEIAKA-----NKO 1717
Qy 1064 VDALKTCACLLKGCRIELAKCIANPSAANVACLOTNNRPD-ETECQ-----IKCGDL 1116
```

```
Db 1718 IESTOKLI-----IPTIKNLISPFKANDLEBGIDTNKMLGKYNTMNNIYEETKSYDL 1770
Qy 1117 -----FENSVVDFNECAVSRKKCVPRKSDGSEFPADPDSVLVQNF-- 1157
Db 1771 ITHYLETWSKEPITYEQIKNKRKITAQNELLTNKNVNAKAKSYLDDIEANEFDRIVTHFKN 1830
Qy 1158 NTSDFNGKW---YITSGLNPTDFADFCOLHEFTEGDNKLVGNISWRIKTLDGSGFFTRSA 1214
Db 1831 KLNDVNDKFTNEY--SKVNGGFDNISNINNKKSTDENLLNINLQTKEM-----YANIV 1884
Qy 1215 VQKVFODPNQPGVLYNHDNEYLHYQDDWYILSSKIENKPEDYIFVYVYGRNDWDGCGGA 1274
Db 1885 SKKYY-----SYKYEAEINFINIPKAN-----SLNTQIKSSSGI 1919
Qy 1275 VVYTRSSVLPNSIILELEKAASIGRDSFTFIRTDNTCGPPALVERLEKTVEEGERI-- 1332
Db 1920 DLFKNITNI---AILPYDSQKK---DTLTFI-----PSP-----EKTSETYTKISD 1959
Qy 1333 -----IVKEVEEIEEVEEVEKEVGRTEMTLFLORAEFGNELKQDEENFVRELSKEEM 1385
Db 1960 SYNTLLDILKRSOELQKQOQALNLI--FENRLLHDKVQATNELK---DTLSDLKNKKE 2013
Qy 1386 EFLDEIKM---EASEVEKL 1401
Db 2014 QILNKVKLLLLHKSNEINKL 2032
RESULT 10
Q49525 PRELIMINARY; PRT; 1365 AA.
AC Q49525;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE LMPI.
GN LMPI.
OS Mycoplasma hominis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2098;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MH56;
RX MEDLINP=95369882; PubMed=7543881;
RA Jensen L.T., Ladefoged S., Birkelund S., Christiansen G.;
RT "Selection of Mycoplasma hominis PG21 deletion mutants by cultivation
in the presence of monoclonal antibody 552.";
RL Infect. Immun. 63:3336-3347(1995).
DR EMBL; U21962; AAA81013.1; .
SQ SEQUENCE 1365 AA; 154982 MW; 847A04992410867F CRC64;
Query Match 2.1%; Score 156; DB 2; Length 1365;
Best Local Similarity 16.8%; Pred. No. 1.3;
Matches 209; Conservative 187; Mismatches 420; Indels 430; Gaps 52;
Qy 272 TRSAVOTFVQDPLPGALYNHDNEYLHYQDDWYILSSQIENKPPDYIFVYVYGRNDWDG 331
Db 282 TRNQTOEFINT-----NKNNP--NYSE-----LISQTSKRD-----SKNSVTDS 319
Qy 332 YGSGVIYTRSTPLPESII-PILOKA-----AKSVGRDFNNFITDNSCGPEPPLVERLEK 385
Db 320 SNKSIESANTELKQALAKANADKVOADNLAKSIKEQLNNSVSNANT-----LSA 369
Qy 386 TAEAGEKLLIKAEVIEEVEEVEKEVGRTEMTLFLORLEGF---RELQDEENFVREL 441
Db 370 KLTDKDNITQOAKTELEKEVQKADQAIKSNNTASQSAKSSLDKVAETKKLETF----- 425
Qy 442 SKEEKEILNELQMEATEVEKLFGRALPIRKLMAVATHCF--TS 960
Db 426 NKDEAKFNLKQTRNQTOEF-----446
```


QY 502 RFSWGWEDYFGSIVVAKICSSRIPIRYFRKSPRICGGLDSRGLQLFSGHKNLSPAH-S-I 560
Db 447 -----INTKNPNYSSELISQTSKRDSKNSVTDSSNKSDESANTEL 489
QY 561 NONPKNGSGCKFPKDVLMVWEKQGFAKTAIVAFILSVASKADADALKTCTCLKE 620
Db 490 KOALAKAN-----ADKVOA-----DNLAKS 509
QY 621 CRLELAKCINPACANVACLOTCNNRPDETECOIKCGDLFENSVDDEFNECAVSRKCV 680
Db 510 IKEQLNNSVSN-----ANTLSAKUTDKDN-----TIOAQKTE 541
QY 681 PRKSDVGDFPVPDPSVLVQKDFMKDFSGKWFITRGLNPTFDA-----FDCQLHEFTEENK 736
Db 542 LEKE-----VQADQAIKSNNTASMSAKSSILDKAVAEITKLTETFNKDKEA 588
QY 737 LVGNLSWRIRTPDGGFTTSRAVQKFDVDPKPGILYNHNDNEYLLYODDWWYLLSSKVENSP 796
Db 589 KFNEKQ-----TRNOIOEFINTNK-----NNPN-----YSELISQTSKRDS-- 626
QY 797 EDYIFVYKGRNDWDGYSVLYTRSAVLPESTIP-----ELQTAQAKVGRDFTFIK 850
Db 627 -----KNSVTDSSNKSDESANTELKQALAKANADKQVADNLAKSIKQALNNSVS 676
QY 851 TDNTCGPEPLVERLEKKEVEGERTIIKEVEIEEVEEK-----VRDKEVTLP 898
Db 677 NANT-----LSAKLTKDNTIOQAQTEKELEKQANQAIKSNNTASMSAKSSLD 726
QY 899 SKLFGFKELQ---RDENFTRLSKEEMVDGLKMEATEVEKLFGRALPIRKLMAVAT 955
Db 727 AKVAETTKLTETFNKDKAEKAKFNQTRNOI-----OEFINTNK----- 765
QY 956 HCFSPCHDRIRFFSDGIGRLGTRKINGTELLKLTPPIQAD-----LRTTGGRS 1009
Db 766 ---NNPNYSEL-----ISQTSKRDSKNSVTDSSNKSDESANTELKQALNTAKAKK 814
QY 1010 S-----RPLSAFRSGFGKIGFDIVPL-----PSKNEL-KELTAPLILLVLG 1049
Db 815 SSIDNELRPL---KNDLOSKIEEFPIRNTNFSWISSKLETTKNKLABELT----- 862
QY 1050 VLACALIVPSADV-----DALTKACALLKGCRIELAKCIANPACANVACLOTCN 1101
Db 863 -----KADAIKNPNSSSKQALKDSSQVQKLGNELKTITE-----EFGKVETKN 907
QY 1102 NRPEDETEQIKCGDLFENSVDDEFNECAVSRKCVPRKSDIGFAPP-----PSVLVQ 1155
Db 908 SNIGYRLFKLAQAEQFNNSVDKLNNAWEEKOTILLSKKQKLGNSQTKDYLTQLSTEMSTQ 967
QY 1156 NFNISD--FNGKWTYISGLNPTFDAPDCOLHEFHTEGDKLVGNISWRIKTLDSGFFTR- 1212
Db 968 ESTIKKIVINQIAHRRNLN-----SQTRLEAD-KLIAN-----MKRGYGDV 1009
QY 1213 --SAVQKFDVDPNPGVLYNHNDNEYLLHYQDDW-----YILSSKIENKPEDYIFVYGRN 1265
Db 1010 GIESLOKQ-QDLMDSDVLSVDDSD---LKDDFNKALRVLVGDYTKNPPVSSWFINKRNS 1064
QY 1266 -DAMDGGYAVYTRSVLPNSI-----IPELEKAASIGR----- 1300
Db 1065 IENTONRLNLIIVRENEILLDKADLDKRAEKTIFKVDENINSLOQARLQOETILNKN 1124
QY 1301 DFSTFIR-----TDNTCGPEPALVERIEKTVEEGERIIVKE-----VEEIEEVEKEV 1348
Db 1125 DLSNFTLNHOKNQBTAKDITPKISLENKLNKLINEINQVLLPIIKKAVYSKISEIEKN-KKEL 1183
QY 1349 EKVGRTEMTLFORLAEGFNLKQDEENFVRELSKEEMEFLDEIKME 1394
Db 1184 EDIIRSNYLWE-----KVEINKYISELTNKQVELRSNINFE 1220

RESULT 11
Q90337
ID Q90337
AC Q90337;

DT 01-NOV-1996 (TREMREL. 01, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
DE MYOSIN HEAVY CHAIN.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprininae; Cyprinus.
OX NCBI_TaxID=7962;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FAST SKELETAL MUSCLE;
RX MEDLINE=97352533; PubMed=9208928;
RA Hirayama Y., Watabe S.;
RT "Structural differences in the crossbridge head of temperature-associated myosin subfragment-1 isoforms from carp fast skeletal muscle.";
RL Eur. J. Biochem. 246:380-387(1997).
RN [2]
RP SEQUENCE OF 738-1933 FROM N.A.
RC TISSUE=FAST SKELETAL MUSCLE;
RX MEDLINE=97176447; PubMed=9023993;
RA Imai J., Hirayama Y., Kikuchi K., Kakinuma M., Watabe S.;
RT "cDNA cloning of myosin heavy chain isoforms from carp fast skeletal muscle and their gene expression associated with temperature acclimation.";
RL J. Exp. Biol. 200:27-34(1997).
DR EMBL; D89990; BAA22067.1; -;
DR EMBL; D50474; BAA09067.1; -;
DR HSSP; P08799; ILVK.
DR InterPro; IPR000048; -;
DR InterPro; IPR001609; -;
DR InterPro; IPR002928; -;
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; -; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
KW Myosin.
SQ SEQUENCE 1933 AA; 221092 MW; CDFOCBAA75530F5 CRC64;

Query Match 2.1%; Score 156; DB 13; Length 1933;
Best Local Similarity 17.0%; Pred. No. 2.1;
Matches 203; Conservative 181; Mismatches 454; Indels 356; Gaps 42;

QY 379 LVERLEKTAEBEGKLLIKEAVEIEEVEEVEKEVDRDTEMTLFORLLEGFKE-----L 430
Db 606 VVQLYQSKALKVLLALLYAVPEAEAGKGGKGGKGS-----FQTVSAVFNENLKLMTNL 661
QY 431 QQDEENFVRELSKEEKE-----ILNELQWE-ATEVEKLFGRALPIR----- 470
Db 662 RSTHPHFVRLCLIPNESKTPGLMENFLVLIHQLCNGVVEGIRICTKGFPRIHYGDFKQRY 721
QY 471 -KLRLMALAPHNSFLANHETIKYVVGSKLPGHFRFSWGWEDYF---GSIIVAKICSSRRIP 526
Db 722 KVLNASVIPESQGFIDNKKATEKLLGSDVDHNQYKFGHTKVFVKAGLGLTLEEMDEKLS 781
QY 527 RYFRKSPRICGLDSRGLQLFSGHKNLSPAH-SINONVPKNGSGCKFPKDVLMVWEK 586
Db 782 HLVTMTQALARGVYMR--KEFVKMMERREAIYSIQYNIRSFNVNKHWP---NMKVVFYKIK 836
QY 587 QFAKTAIVAFILSVASKADADALKTCTCLKECELELAKCINPACANVACLOTCNN 646
Db 837 PLLKTAESE---KEMASMKENFEKMKEDLTAKLAKKKELEKMSVLVQBNKLLLOVTSE 893
QY 647 RPDETECOIKCGDLFENS-----VDEFECAVSRKK-----CVPRKSDV 686
Db 894 SENLSDAEERCEGLIKSIQLEGKLEKLTETTERLEDEEINELTAKRKLKLEDECSSELKKDI 953
QY 687 GDFVPDPDPSVLVQKDFMKDQSGKWFITRGLNPTFDAPDCOLHEFHTENKLVNLSWRIR 746

RA RA RA RA RA RA RA RA RA RA RA RA RA RA RT RL DR DR DR DR DR KW SQ	<p>Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E., Ketchum K.A., Dodson R.J., Winn M., Hickey E.K., Peterson J.D., Richardson D.L., Kerlavage A.R., Graham D.E., Kyriades N.C., Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S., Kirness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B., Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L., Overbeck R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T., Cotton M.B., Spriggs T., Artiach P., Kalne B.P., Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;</p> <p>The complete genome sequence of the hyperthermophilic, sulphate- reducing archaeon Archaeoglobus fulgidus.";</p>		Query Match Best Local Similarity 2.1%; Score 155.5; DB 1; Length 886; Matches 137; Conservative 85; Mismatches 221; Indels 221; Gaps
QY	797 EDIFYVVKKGRNDWDGYSGV--LYTSAVLPSIIPELOTAQAQVGROFNFTIKTD-N 853	: :	
Dd	152 EDY-----ENAWNLGAVMRLERKERKEFLSOEQIQRKQEEK -----KAIE 197	: :	
QY	854 TCGEPPPLVERLEKKVEEGERTI--IKEVEEIEEVVKVDKVETLFSKLFGFKELQR 910	: :	
Dd	198 RISETKIESTRESLKLSSEVRNLESRLKELEHKSRLSLRKQESSVLQEVS ----R 249	:	
QY	911 DEENFEELSKEEMDV---LDGLKMATEVSKLGFRALPIKLMVAATHCTSPCHDIR 967	: :	
Dd	250 GLSEKURELEKQLKEVVEDILEKKAKEVEL-----KPAAERY - 289	: :	
QY	968 FFSDDDGIGRLGITRRKRNGTFLLKILPIQSA-DLRTTGGRSSPLSAFRSGFSKGIF 1025	- - - - -	
Dd	290 -----SILEKLSEINQALRDYEKREGDLTRAAGIQAOLKAE 329	: :	
QY	1026 DIVPLPSKNELKETAPL-----LLKLVGLACAFILVPSSADAVDAKTIC 1070	: :	
Dd	330 D-----NSKLEEITRIELELEREFEXSHRLE-----TLPKMDRMQGIIK -- 372	: :	
QY	1071 ACILKGCRIELAKIANACAANVACLQTNNREDTECOICKDGLFENSVDVFNECAV 1130	- - - - -	
Dd	373 -----AKLEEKMLTPDKVE----KMYDLLSRKAKEE-KETE 404	: :	
QY	1131 SRKCKVPRKSDLGPDPDPVSVQNPFNISDFNGKWYTISGLNPSTDPAFCQHFFHTGEG 1190	: :	
Dd	405 KLKLTIAXKSSLTRGAOLKAAVE-----LKSABERTCPVGRELDEHRKN 451	-- - - -	
QY	1191 -----DNKLVGNISWRILTKDSGFFTSAVOXFPQPNPGVLYN-- 1230	: :	
Dd	452 IMAYETREMKLIAELAKADEIKIERUEKYEKALEKOETIVKYRWQMVDELKALENEL 511	:: :::: :	
QY	1231 --HDNEYLUHYODDYILLCSKENPEDYIPFYVYRGNRDAMDYGGAHVYTVRSVLPNSII 1288	:	
Dd	512 SSHDAE-----KLSASESB-----YRVKNERLDGLRGOOKTILLSA--SRI 550	: :	
QY	1289 PELEKAASKIGRDEFSTIRDTNTCGEPVALPERIAKTYEEGERIIVEKEELEBEVEKEV 1348	: :	
Dd	551 KELASSUREI-----EHAL-KNVESSELHKRIREFGFESLFERELEV 593	: :	
QY	1349 EKVGRTMTLFOELAEGFN---ELKDQDENFVREISKSEMPEFDIEIKMEASEVEKLEFGKA 1405	: :	
Dd	594 -----QSURPYPNKWLKEDAESRLESCLKRB-KLEDISEISAIAKLEANCKA 641	: :	
QY	1406 LPID 1409		

```

Db 642 EIR 645

RESULT 13
Q95722 PRELIMINARY; PRT: 1055 AA.
AC Q95722;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE SMC-LIKE PROTEIN.
GN MIM.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsiis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=99380167; PubMed=10449416;
RA Mengiste T., Revenkova E., Bechtold N., Paszkowski J.;
RT "An SMC-like protein is required for efficient homologous
RT recombination in arabidopsis.";
RL EMBO J. 18:4505-4512(1999).
DR ENBL; AFI20933; RAD54770.1;
DR ENBL; AFI20932; RAD54769.1;
DR InterPro: IPR003439;
SQ SEQUENCE 1055 AA; 121349 MW; E9F0C0427FB602E4 CRC64;

Query Match 2.0%; Score 151; DB 10; Length 1055;
Best Local Similarity 19.3%; Pred. No. 1.8;
Matches 182; Conservative 119; Mismatches 306; Indels 334; Gaps 43;

Qy 128 ALKTCACLLKECRIE-----LAKCIANPSCAANVACIOTCNRRPDETEQCQKCGDLLEN 181
Db 57 AILTALCIAFGCRARGTQRAATLKDFIKTGCYAVVQVEMKNSGEDAFKSEIYGVIIIE 116
Qy 182 SVVDQFNECAVSR----KKCPKSDVGEPVPDRNAVQVQNNM-----KDFSGK 227
Db 117 RRITESATATVLDKGLKSVNSRDELRE-----LVEHNDIVENPCVVSQDKAGS 168
Qy 228 WITSGLNPTFDAPDCQLHEFHENDKLVGNLTWRIKTLDDGFFTRKSAVQTFVQDPDLP 287
Db 169 SYI-----LECK-----GN-----SSFLNLIQ-----QVNDLIQ 194
Qy 288 ALYNDHNEFLHYQDDWILTSQIEN--KPDYIFVYGRNDADWDGYSVIYTRSPILP 345
Db 195 SIYELHTKATAIVD-----ELENTIKPEKEISELRGK-----IRNMEQV 234
Qy 346 ESIIIPNLQKAAK-----SVGRDNFNFTTNSCGPEPPLVERLEKTAEGEKLLIKE 397
Db 235 EEIAQRLOQLKLLAWSWVDVGRQ-----LQEQTEKIVKLKERIPTCQ 278
Qy 398 AVEIEEEVEKEVKVRT-----EMTLFORLLEGKE-----LQDD 433
Db 279 A-KIDWELGK-VESILRDTLTKKKAQVACLMDSTAMKREIESHQSAKTAVREKIALQEE 336
Qy 434 -----EENFVRELSKEKEILNELOMEATEVEKLFGRALP 468
Db 337 FNHKONYOKTKDVRRLERQVGOINQOTMKNTOAQEISEEKLKYLEREVEK----- 389
Qy 469 IRKLRMALAPHSN-----FLANHETIKYVYVGSKLPGHKRFSGWEDY 510
Db 390 VETLSRLKEEENCFLEKAFEGKMKMEHIEDMIKNHQRQRFITSNINDLKKHQTNRKVA 449
Qy 511 FGSIVVAKICSS--RRIPYFRKSPRICGLDSRGLQLFSGHKHNLSPAHNSINQNVKGS 569
Db 450 FGGRVINLLOAIERNHRRFRKPP-----IGPGS----- 479
Qy 570 GCKFPKDVAMVWEKVGOFAKTAIVAIFILSVASKADAVDAKTCCTCLLKECRELAACI 629

```

```

Db 480 -----HVTLVGNKNWASSVEQALGT--LLNAFIVTDHKDSL----- 513
Qy 630 SNPACAAVACIOTCNRRPDETEQCQIKCGDLFENSVDNECAVS--RKCVPRKSDVGD 688
Db 514 -----TLGCANEANYRNLI-----IYDFSRPLNPRHVPQTEHPTI 554
Qy 689 FPVPDPSVLVOKFMKDFSGKWFITRGLNPTF-----DAFDCQLHEPTEENKLV--- 738
Db 555 FSVIDS-----NPTFLNVLVDQSGVERQVLAENYEEKAVAFG 593
Qy 739 ---GNLSWRIRTPDG--GFTSRSAVQKFQD--PKYPGLIYNHDNEYLLYQDDWYLSKVE 793
Db 594 KRLSLNK-EVYTLGYKMFPRGPVQTTPLPSRRPSRLCASFDQI---KDLTEASKEQ 649
Qy 794 NSPEYIFVYKGRNDAMDYGGSVLYTRSAVLPEIPIPELOTAQKVGDRFNTFIKTDN 853
Db 650 NE---INQMRKRKEAENLELEKLVKRLKRSQAEKVLTTKELEMDLK-----N 699
Qy 854 TCGPEPPLVERL-EKKVBEGETIIEVEEIEEE---VEKVRD--KEVTL----FSKLFE 903
Db 700 TVAAE---IESLPSSSVNELQREIMKDLDEIDEKEAFLEKQNCLEKAEKANKLITALE 756
Qy 904 GKELQDRE---ENFLRELSKEEMDVLDGLKMEATEVEKL 940
Db 757 NMRESAKGEIDAFEEAENELKIEKD-----LQSAEAEKI 791

RESULT 14
Q25662 PRELIMINARY; PRT: 1939 AA.
AC Q25662;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)
DE REPEAT ORGANELLAR PROTEIN.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=96V;
RA Werner E.B., Taylor W.R., Holder A.A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U43145; AAC63403.1;
SQ SEQUENCE 1939 AA; 229001 MW; B36E462001C6F22F CRC64;

Query Match 2.0%; Score 151; DB 5; Length 1939;
Best Local Similarity 17.5%; Pred. No. 4.1;
Matches 189; Conservative 167; Mismatches 299; Indels 422; Gaps 50;

Qy 380 VERLEKTAEGEKLLIKEAVEIEEEVEKEVKYRDTMTLFORLLEGKELQD-----EEN 436
Db 95 VKEYEVIEELEKEL---KLEKEQINKYEKELNEKSEFIKQKMLLEKELNINLEN 151
Qy 437 FVRE-----LSKEKEILNELOME---ATEVEKLFGRALPIRLKRMALAPHSNLANHE 487
Db 152 KINKEIITLKREEK--LNDIESEYIEKNKEKEL---NYEVTNKMSL-----D 196
Qy 488 TKYVYVGSKLPGHFRFSGWEDYFGSIVVAKICSSRRIPYFRKSPRICGLDSRGLQLF 547
Db 197 KLTCEVQEKKNLEKIN-----KKVIEKENNRELKEFMEKKEKNEIISLDG----- 242
Qy 548 SHGRHNLSPAHNSINQNVKNSGCKFPKDVAMVWEKVGOFAKTAIVAIFILSVASKADA 607
Db 243 -----TIIN-----DKNAYEKLEI-----SFEKKRM 264
Qy 608 VDALKTCCTCLLKECRELAACISNPACAAVACIOTCNRRPDETEQCQKCGDLFENSVD 667
Db 265 IEMLD-----SKLIEKEENFAN-----KOAKLE 287
Qy 668 EFNCAVSRRKCVPRKSDVGDFFPVPDPSVLVQKFDKMDKFSCKWFTTRGLNPTDAFDCQL 727

```

```

Db 288 KENIIIEKLDIESRE-||| |KDFSK-----EEKFASMENEL 321
QY 728 HEFTEENKLVNLS-WRIKTPDGFRTSAVQKFDQDQKPYGILYNHNDNEVLLYQDDWY 786
Db 322 NTLKSDLSKACOMEYKLEIKD---LSQSLVEKERE-----IFEIKNEY---DD-- 365
QY 787 ILSKVENSPEDYIFVYKGRNDAMDYGGSVLYTRSAVLPSIIPESIIPELQTAQKVGDRDN 846
Db 366 ----KINMKKLSI---ND-KGIDNTVLHSEEEKI-NKLLKEKETELNEIHKYIN 413
QY 847 TFIKTDNTCGPEPPLVERLEKKEVGEERTIIKEVEIEEVEKVR-----DKEVT 896
Db 414 LEIE-----TIKNELNEKEEELKKNKAHTVEVTNLTKIEK 449
QY 897 LFSKLFEGFKELQDE---ENFLRELSKEEMDVLGDKMEATEVEKELFGRALPIRLKMA 952
Db 450 LLEKKTDEAKEGHKNELNQLSKLNEK---DNINKNTEL----- 490
QY 953 VATHCTSPCHDRIFRFSDDGIGRLGTRKRINTFTLLKILPPIQSADLRTTGRSSRP 1012
Db 491 -----NDKISLNS-----VN-----ILNKDKOTLG----- 512
QY 1013 LSFRSGFSKGFIDVPLPSKNELKELTAPLLLLKLVGLVLAACFLVPSADAVDAKTCAC 1072
Db 513 -----NDIKTLN-----DINNLKN----- 527
QY 1073 LLKGCRIELAKIANPACAANVACLOTNNRPDE-----TECOIKCGDLFENSVD 1123
Db 528 -----EINTSDKNMKMKEDLAMLNEMEGKC-----VVID 558
QY 1124 EF-----NECAVSRKKCVPRK---SDLGFPAPDPSVLQVONENISDFNGKWIYISGLNPTF 1176
Db 559 ETEKKYKEIFMELEKKEKENYADLND---EISILNLSYVRE---KEFI----- 603
QY 1177 DAFDQOLHEFTGDNKLVGNISWRIKTLDSGFFTRSAVOK---FVODPNQPGVLYNHND 1233
Db 604 -----EMKEFY---ENKI-----NLENKPEEKNIYENELNLSRLKYDNEQ 642
QY 1234 EYLHYQDDWYILSSKIENKPEDYIFVYKGRNDAMDYGGAUVYTRSSVLPNSIIPELEK 1293
Db 643 GLIKQIDELNTOQLKTEK---YLOLY-----NDN-----MHMFSICTKIDMYPSEN 687
QY 1294 AAKSGTRFST-FI---RTDNTCGPEPALVERIEKTVBERIEIYKEVEEIEEVEKEVEK 1350
Db 688 IKGSDLVDFVYAYIKRRDESSDAMPDTHK--EMVALEKRAHAAVIAELEKHEKEIAK 745
QY 1351 VGRTEMTLQRLAEGFNE---LKQDEENFVRELSKEEMEFLEIKMEASEVEKLF 1403
Db 746 LQEGHKEVVLRLGEOHKEETIILEEKHKDVVTKLGEQKENIILKEEHKDVVTKLG 802

RESULT 15
Q9HAW4 PRELIMINARY; PRT; 1332 AA.
AC Q9HAW4
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE HU-CLASPIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21000493; PubMed=11090622;
RA Kumagai A., Dunphy W.G.;
RT "Claspin, a novel protein required for the activation of Chk1 during a
RL DNA replication checkpoint response in Xenopus egg extracts.";
DR Mol. Cell 6:839-849(2000).
DR EMBL; AF297866; AAG24515.1;
SQ SEQUENCE 1332-AA; 150175 MW; 4EA38DC16A3936C3 CRC64;

```

```

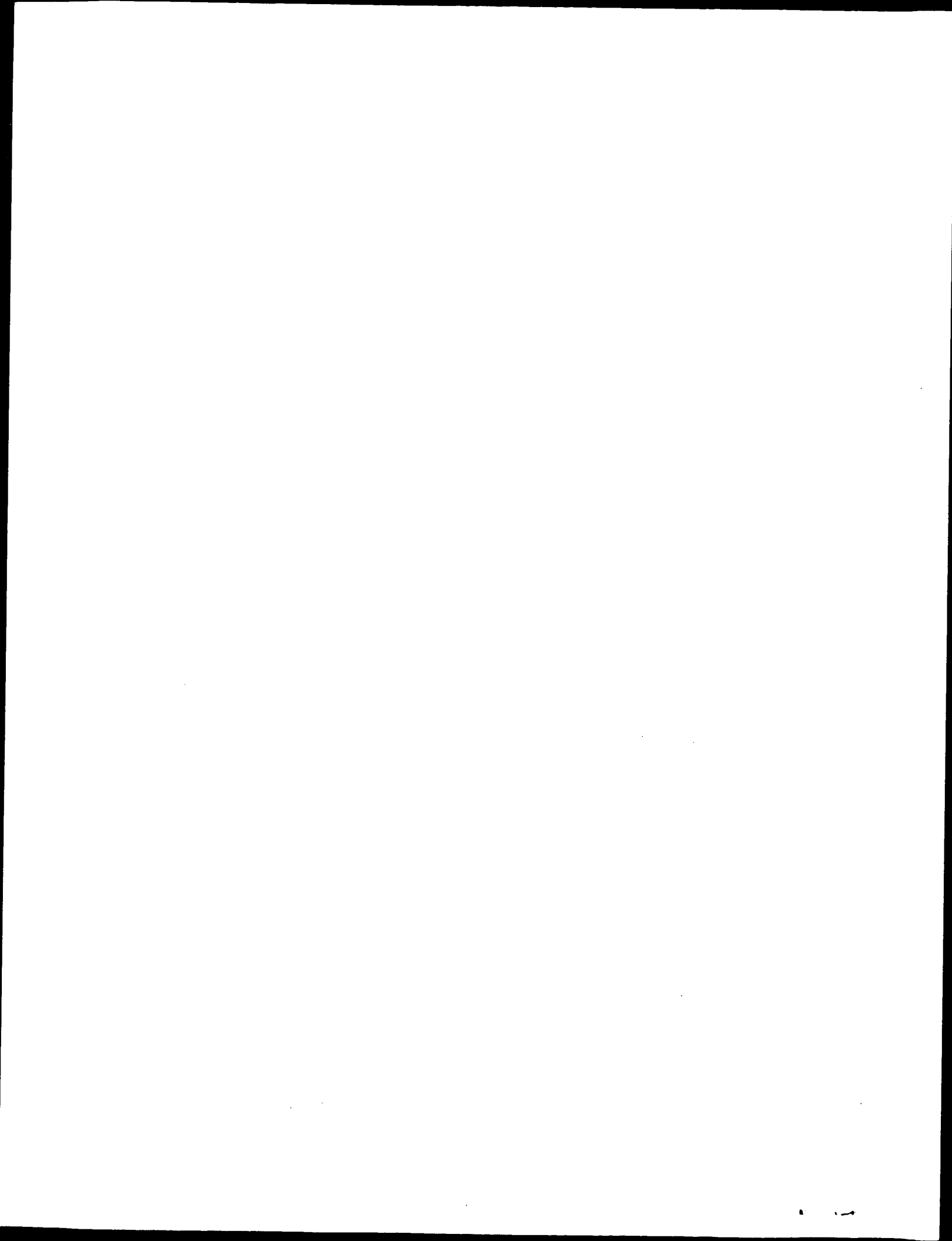
Query Match 2.0%; Score 150.5; DB 4; Length 1332;
Best Local Similarity 19.3%; Pred. No. 2.6;
Matches 255; Conservative 182; Mismatches 405; Indels 477; Gaps 67;

QY 307 SSOIENKPPDYIFVYGRN-----DAMDYGGSVYTRSPPLPESIIIPNLOKA 355
Db 83 SAEENKEN-----LYAGKNTRIKRIYKTADVSDSEYKESLYQEN--LEAQVKPCLELS 135
QY 356 AKS-VGRDFNNFITDNSCGPEPPLVERLEKTAEBEGKLLIKEAVELEEEVEKEVEKVRD 414
Db 136 LOSGNSTDF---TTDRKSSKH--IHDKGTA---GKAVKSRRLKEKE-ERAKMKIRQ 185
QY 415 -----TEMPLFORLLEGGFKELQ-QDEE-----NFVR 439
Db 186 LKKKETKNQEDDVEQFNDSCGLLVKDKLFFTGLEDENNNSPLEDEESLESTRAAVKNKV 245
QY 440 ELSKEKEETLNELQW--EATEVEKILFGRALPIRLKMA--LAPHSNFLANHETIKYIVGS 495
Db 246 KHKKEPSLESCVHSFEGSELK--GTT---RKERKAARLSKEALKQLHSETQRLIRES 300
QY 496 --KLPGHRRFSWGMEDYFGSIVAKICSSRRIPRYFRKSPRICGLDSRGL-----QLFS 548
Db 301 ALNLPYH-----MPENKTIHDFKRRKPRPTCHGNAMALLKSSKYQSS 342
QY 549 HGKHNLSPAHSINQVPGNSGCKFPKDVLMVWEKQGQAKTAIIVAIFILSVASKADAV 608
Db 343 HHIKEIIDTANTENMSDHSKGS-----OTTGAENEVETNA---LPVYSKETQI 389
QY 609 DALKTCTCLLKECRLELAKICISNPACAANVACLOTNNRPDETECOIKCGDLFENSVD 668
Db 390 -----ITGSDSCRKDLVK-----NEELEIQEK----- 412
QY 669 FNECAVSRKKCVPRKSDVGDFPVPDPSVLVQKDFMDKDFCKWIFITRGLNPTFADFPCQLH 728
Db 413 -----QKQDIRSP-GDSSVLOQESNF-----LG 436
QY 729 EFHTEENKLVNLSWRIKTPDG-GFFTRSAVQKRVODP----- 765
Db 437 NNHSEECQVGLVAFEPHAGEGEPQNPETDEKVEPEQONKSSAVGPEKVRRTLDR 496
QY 766 -KYGILYNHNDNEVLLYQDDWYILSSKVENSPEDYIFVYKGRNDAMDYGGSVLYT--- 821
Db 497 LKQGVDSVSKPRIGADEDSFVILEPETNRELEALKQREWKHANPAKPRAGQTVNVVI 556
QY 822 -----RSAVLPESIIP-----ELQTAQKVG-----RDF----- 845
Db 557 VKDMGTGKEELADADVVPVTLAPKLDGASHTKPGKQLVLKAKLQEAAMKLRFRFEOR 616
QY 846 NTFIKTDNTCGPEPPLVERLEKKEVEGERTIIKEVEIEEVEKVRDKRVTFLSKLFEGF 905
Db 617 QALFKLDNEDGSE-----EEEEEEEMTDESE-EDGEKVEK-EEKEELEEE-EGK 665
QY 906 KELQDEEN-----FURELSKEEMDVLGDKME-----ATEVEKILFGRALPIRLKMA 954
Db 666 EEEEEEGNOETAEFL--LSSEIETKDEKMDKENNDGSSSEIGRAVGFLSVPKSLSDS 723
QY 955 THCTFTSPCHDRIFR-----FSSDDGIGR-----LGITRKRINGTFIL--KIL 994
Db 724 TLLFKDSSSKMGYSPTSEKSETDENSQPSKLDDEDDSCSLLTKESSHNSFELIGSTI 783
QY 995 PPIQSADLRTTGRSSRPLS-AFRSGFSKGIF-DIVPLPSKN--ELKELTAPLLKLVG 1049
Db 784 PSYQPCNROTGRGTSFFPTAGGRSP-SPGLFRASLVSSASKSGKSEPSLP----- 835
QY 1050 VLACAFILVPSADAVDAKTCACLLKGCRIELAKIANPACAANVA-CLQTCNNRPDETE 1108
Db 836 -----IEDSQDLYNASPEPKTLFLG-----AGDPQFCLE-----DDTQ 868
QY 1109 CQIKCGDLF-----ENSV-----VDFNECAVSR-----KKCV 1137
Db 869 SOLLDADGFLVNRHNRNOYQALKPRPLASMDENAMDMDELIDCTGKFTSOAEKILP 928

```

1138	RKSD	-----LGEFAPDPSPVLVONFISDNFK-----WITSGL	1172
	QY		
929	RKSD	KKENMEELLNLCGKFTSDASTPAS-----SELNKQEKSSGDPMEELALCSGS	984
	Dbb		
1173	NPTF	DAFCQLHFHTGDKNLKVNISWRIKTLDSGFFTRSAVQKFPQDPNPQGVLYNHD	1232
	QY		
985	FTTK	EEDEEERF---GDFRLYSN-----DNEF---DSDEHSDSGNDLALEDHE	1030
	Dbb		
1233	NEYL	HYQDOWYL---SSKIENKPEYIFVYVYGRNDAMDYGGVYVTRSSVLPNSIPE	1290
	QY		
1031	DD	-----DEELLKRSEKLK-----RQMLRYLEDE	1057
	Dbb		
1291	LEKAKS	IGRDFSTFIRDTNCTGPEPALVERIEKTVVEGERIIVKEVEIEEVEKEVEK	1350
	QY		
1058	AEVSGS	VGSE-----DEYDGE-----IDYEEDVIDELPSDEELQSOIKK	1100
	Dbb		
1351	V	-----GRTEWTLFOR---LAGFNLKQDENFVRELSEKMEFLEIDKWEASEVE	1399
	QY		
1101	IHKMTL	DDDKRLRLQYERYLADG---DLASDGGPRMKRFWKND-----DASQMD	1150
	Dbb		

Search completed: November 6, 2001, 05:13:20
Job time: 574 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2001, 03:04:01 ; Search time 6769.47 Seconds
(without alignments)
3553.062 Million cell updates/sec

Title: US-09-075-375A-5
Perfect score: 1555
Sequence: 1 ccacgcgtccgctgtgtgt.....ggataaaaaaaaaaaaaa 1555

Scoring table: IDENTITY_NUC
Gapop: 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba1.*
2: gb_ba2.*
3: gb_ba3.*
4: gb_in1.*
5: gb_in2.*
6: gb_in3.*
7: gb_om.*
8: gb_ov.*
9: gb_pat1.*
10: gb_pat2.*
11: gb_ph.*
12: gb_pl1.*
13: gb_pl2.*
14: gb_pl3.*
15: gb_pl4.*
16: em_ba1.*
17: em_ba2.*
18: em_fun.*
19: em_htgo_hum.*
20: em_htgo_inv.*
21: em_htgo_rod.*
22: em_htg_hum1.*
23: em_htg_hum2.*
24: em_htg_hum3.*
25: em_htg_hum4.*
26: em_htg_hum5.*
27: em_htg_hum6.*
28: em_htg_hum7.*
29: em_htg_hum8.*
30: em_htg_inv1.*
31: em_htg_inv2.*
32: em_htg_other.*
33: em_htg_rod.*
34: em_hum1.*
35: em_hum2.*
36: em_hum3.*
37: em_hum4.*
38: em_hum5.*
39: em_hum6.*
40: em_hum7.*
41: em_in.*
42: em_om.*
43: em_or.*

44: em_ov.*
45: em_pat.*
46: em_ph.*
47: em_pl.*
48: em_ro.*
49: em_sts.*
50: em_sy.*
51: em_un.*
52: em_v1.*
53: gb_sts1.*
54: gb_sts2.*
55: gb_sts3.*
56: gb_sy.*
57: gb_un.*
58: gb_v1.*
59: gb_v12.*
60: gb_htg1.*
61: gb_htg2.*
62: gb_htg3.*
63: gb_htg4.*
64: gb_htg5.*
65: gb_htg6.*
66: gb_htg7.*
67: gb_htg8.*
68: gb_htg9.*
69: gb_htg10.*
70: gb_htg11.*
71: gb_htg12.*
72: gb_htg13.*
73: gb_htg14.*
74: gb_htg15.*
75: gb_htg16.*
76: gb_htg17.*
77: gb_htg18.*
78: gb_htg19.*
79: gb_htg20.*
80: gb_htg21.*
81: gb_htg22.*
82: gb_htg23.*
83: gb_htg24.*
84: gb_htg25.*
85: gb_pr1.*
86: gb_pr2.*
87: gb_pr3.*
88: gb_pr4.*
89: gb_pr5.*
90: gb_pr6.*
91: gb_pr7.*
92: gb_pr8.*
93: gb_pr9.*
94: gb_rol.*
95: gb_rod.*
96: gb_in4.*
97: gb_pr10.*
98: em_ba3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1555	100.0	1555	14	ATU44133
2	958	61.6	133843	12	AC003981
3	922	59.3	2884	14	ATU39452
4	624.2	40.1	1589	14	NTU34817
5	611.8	39.3	1760	14	LSU31462
6	601.4	38.7	1686	15	SOL250433
7	444.8	28.6	1173	13	AF288196
8	330.4	21.2	89479	12	AC006932
					U44133 Arabidopsis
					AC003981 Genomic s
					U39452 Arabidopsis
					U34817 Nicotiana t
					U31462 Lactuca sat
					AJ250433 Spinacia
					AF288196 Oryza sat
					AC006932 Genomic s

```

c 9 92.4 5.9 7218 10 I66494
c 10 56 3.6 147243 84 HSAJ9612
c 11 55.6 3.6 39987 4 AC009601
c 12 55.6 3.6 42160 4 AC005804
c 13 55 3.5 162896 67 AC023131
c 14 55 3.5 178617 66 AC021529
c 15 54.8 3.5 2813 14 CHOOORE5
c 16 54.8 3.5 185931 76 AC083812
c 17 54.2 3.5 100000 91 AP000067
c 18 54.2 3.5 103908 77 AC073799
c 19 54 3.5 166335 75 AC073799
c 20 54 3.5 199036 80 AL357376
c 21 53.2 3.4 200000 60 AC004670
c 22 53 3.4 125170 76 AC079627
c 23 52.8 3.4 169931 86 AC005822
c 24 52.4 3.4 118632 89 AL133380
c 25 52.4 3.4 204099 69 AC025669
c 26 52.4 3.4 212671 77 AC084849
c 27 52.2 3.4 80017 64 AC016368
c 28 52.2 3.4 151802 79 AL354867
c 29 52.2 3.4 194125 70 AC026383
c 30 52.2 3.4 198983 82 AL589874
c 31 52.2 3.4 205365 78 AC090437
c 32 52.2 3.4 221676 82 AL589876
c 33 52 3.3 4238 95 RNCNG41
c 34 52 3.3 157921 71 AC027726
c 35 52 3.3 166871 61 AC010647
c 36 52 3.3 176364 61 AC009785
c 37 51.8 3.3 1228 7 MRU59897
c 38 51.6 3.3 218846 81 AL589701
c 39 51.4 3.3 223000 77 AC087556
c 40 51.2 3.3 110000 76 AC079517_1
c 41 51.2 3.3 10000 76 AC079517_2
c 42 51.2 3.3 169434 90 AL359205
c 43 51.2 3.3 171588 67 AC023198
c 44 51.2 3.3 198692 78 AC090530
c 45 51 3.3 58484 89 AL137122

```

ALIGNMENTS

```

RESULT 1
LOCUS ATU44133 1555 bp mRNA PLN 15-JUN-1998
DEFINITION Arabidopsis thaliana violaxanthin de-epoxidase precursor (AVDEL)
ACCESSION U44133
VERSION U44133.1 GI:1465734
KEYWORDS thale cress.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1555)
AUTHORS Bugos,R.C., Hieber,A.D. and Yamamoto,H.Y.
TITLE xanthophyll cycle enzymes are members of the lipocalin family, the
first identified from plants
J Biol. Chem. 273 (25), 15321-15324 (1998)
MEDLINE 9628256
REFERENCE 2 (bases 1 to 1555)
AUTHORS Bugos,R.C. and Yamamoto,H.Y.
TITLE Direct Submission
JOURNAL Submitted (02-JAN-1996) Robert C. Bugos, Plant Molecular
Physiology, University of Hawaii, 3190 Maile Way, Honolulu, HI
96822, USA
FEATURES
source
1. .1555
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
1. .1555
gene

```

```

CDS
/gene="AVDEL"
45..1433
/codon_start=1
/product="violaxanthin de-epoxidase precursor"
/protein_id="AAC50032.1"
/db_xref="GI:1465735"
/translation="MAVATHGCTSPCHDIRFFSSDDGIGRLGTRKINGTFLKIL
PPIQSADLRITGGSSRPLSAFRSFGKIEDIVPLPSKNEKELTAPLLKLVGLA
CAFLIVFSDADVDAKTCACLLKGRICAKIANPACAAVACLOTCNNRPNDETECQ
IKGDLFFNSVVDFNEFCAPKRSKLDGFFPAPDPSLVNFINISDFNGKWI
TSLNPTDFADQDLHEFTEDKNKLVNISRKILDSGFFTSVAKVQVQDNPQG
VLYNHDYLLHYQDDWYILSSKIKENKPEDYIFVYVRNDAMDGYGGAIVYTRSVLP
NSIIPELKKAASKIGRDFSTFIRTDNCGPPALVTERKTEVEGERIIVKEVEIEE
EVEKEVEKVGKTEMTLFLAEGFNELAQDENFVRELKEMEFLDEIKMEASEVEK
LFGKALPRKVR"
transit_peptide 45..383
mat_peptide 384..1430
/gene="AVDEL"
/gene="AVDEL"
/product="violaxanthin de-epoxidase"
/db_xref="GI:1465735"
BASE COUNT 485 a 285 c 376 g 409 t
ORIGIN
Query Match 100.0%; Score 1555; DB 14; Length 1555;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 caacgcctcgcgccttgctgctgggaagattagatagtggaagaatggcagtagctaac 60
DB 1 CCACGCGTCGCGCTTGGTGGGGAAGATTAGATAGTGTGAAGAATGGCAGTAGTACAC 60
QY 61 atgtttcaactcactctgactgacccgtattcgattttctcaagtgtgatgtattg 120
DB 61 ATGTTTCACTTACCTTGCATGACCGTATTCGATTTTCTCAAGTATGATGATGATG 120
QY 121 gttagcttggtcattacaagaagagatgcaatggcactttcttgcctcaagattttac 180
DB 121 GTAGGCTTGGCATTACAGAAAGAGAGATCAATGGCCTTCTTGTCTAAGATTTTACCTC 180
QY 181 caatcaagaagtctgactccagacaactgggtggagatccctcacgtctttattctcat 240
DB 181 CAATCAAGAAGTCTGATCTCAGAACAACTGGTGGAGATCTCAGTCTCTTATCTGAT 240
QY 241 tcagtcaggattctctaaaggggattattgacattgtgacctaccatcaagaatgagc 300
DB 241 TCAGTTCAGGATTTCTTAAGGGGATATTGACATTTGCCATTACCATCAAGAATGAGC 300
QY 301 tgaagagctgacgcctcgcctgttgcctaaactcgtggtgttttagcttgcgcgtcc 360
DB 301 TGAAGAGCTGACCGCTCCGCTGTGCTTAAACTCGTGGGTGTTTAGCTTGGCGGTTC 360
QY 361 ttattgttcacttcagatgcagttgatgcacttaaaacttgatgcattgtattgaagg 420
DB 361 TTATTGTTCCATCTCAGATGCAGTTGATGCATTAAACTTTGTCATGCTTATTGAAG 420
QY 421 gatcagagatagaactcgaagtgcatcgcacacctcgcctgtgacccaatgtcgcgt 480
DB 421 GATCAGGATAGAAGTTCGCAAGTGCATTGGCAACCTGCGCTGTGCAGCCAATGTCGG 480
QY 481 gcttcagacctgcaataaccgctccagatgaacccagatgacagattaaatgtgggagc 540
DB 481 GCCTTCAGACCTGCATTAACCGTCCAGATGAACCGAGTGCAGATTAAATGTGGGATC 540
QY 541 tttttgagaacagtgcttggatgagttcaacagagtgctgtgtgtcgagaaaaagtgtg 600
DB 541 TGTTTGAGAACAGTGTGTTGATGAGTTCAACAGAGTGTGCTGTGTCGAGAAAAAGTGTG 600
QY 601 ttcttagaaaaatctgactcggaaaaatttctgccagaccctctgttctttacaga 660
DB 601 TTCTTAGAAAAATCTGATCTCGAGAAATTTCTGCCCCAGACCCCTTCTGTTCTTTACAGA 660

```


285 ccatcaagaatgagctgaagaagctgaacgcctccgctgttgctaaaaactcgtggtgtt 344
 5703 CCATCAAGAATGAGCTGAAGAAGCTGACCGCTCCGCTGTGTGTAACACTCGTGGGTGT 5762
 345 ttagcttgcggtccctctattgttccatctgcagatgcagttgatgcacttaaaacttgt 404
 5763 TTAGCTTGCGGTCCCTATTGTTCATCTGCAGATGCAGTGTGATGCACCTTAAACTGT 5822
 405 gcatgctattgaaggatgc----- 425
 5823 GCATGCTATTGAAGGATGACGTAACCTTTCCCTTCGTTGGATTCTGCATATATAGTGT 5882
 426 ----- 425
 5883 CCCCCCAAAACCAGTTTATTTTCATTGTGATGTGATGCTATAAGTGTGTTTGAATT 5942
 426 ----- 425
 5943 GTATAGTCCATCTCTCAATGTTGGAATGTTAAAGCTAGATGAAGTACTGTGTCT 6002
 426 ----- 425
 6003 GTCTATGATAGAGCTGAAATCAGTTGATGATATCTTTCTGTATTTTTCGCGTCAATTTA 6062
 426 ----- agatagaactcgc 439
 6063 TATATTTCCCTCCTGTCTGACCATATGGAAGCCCAATTATCTGTATTAGGATAGAACTCGC 6122
 440 aaagtgcattccaaacctgcctgcagccaatgtcgcgtgccttcagacctgcaataa 499
 6123 AAAGTGCATTCGCAACCCCTGCTGTGAGCCAAATGTGCGTGCCTTCAGACCTGCATATA 6182
 500 ccgtccagatgaaccaggatgc----- 521
 6183 CGGTCCAGATGAACCCGAGTCCAGGTATATAATTTATTTTGATATATCAATTTTATGAGGC 6242
 522 ----- 521
 6243 AGAGATATTGCTTGCTTAATGAAGAAGAACCAATGATTTGTTTAAATATGTGCTTATA 6302
 522 -----cagattaaatggggatctgtttgagaaacagtggttgatgagttcaacag 575
 6303 TTATATCAGATTAAATGTGGGATCTGTTCGAAACAGTGTGTGTGATGAGTTCACAGAG 6362
 576 tgtctgtcgcgagaaaaagtgttctcctagaaaaatctgactcgcgagaatttctcgtcc 635
 6363 TGTGCTGTCTGAGAAAAAAGTGTCTCTAGAAAAATCTGATCTCGGAGAAATTTCTCTGCC 6422
 695 ccagacctctgtcttctgtacagaactccaacatctcgcgactttaacggaagtgtac 695
 6423 CCAGACCCCTTCGTCTTGTACAGAACTTCACAFCTCGGACTTTAACGGGAAGTGGTAC 6482
 696 attacaagtggcttgaatccaaacctttgatgccttcgactccagctgcagttatccac 755
 6483 ATTACAAGTGGCTTGAATCCAACTTTGATGCTTCGACTCCAGCTGCATGAGTTCAC 6542
 756 acagaagtgacacaacagcttgttggaacatctctcttgagaaataaagacctagacagt 815
 6543 ACAGAAGGTGACACAACAGCTTGTGGAAACATCTCTTCGAGAATAAAGACCCCTAGACAGT 6602
 816 gcatcttactagtcagcgcgtacaaaaattcgtgcaagatccctaaaccaacctgggttt 875
 6603 GGATTTCTTTACTAGTGCACCGCTACAAAAATTCGTGCAAGATCTTAACCACTGGTGT 6662
 876 ctctacaatcatgacaacaggtacacctcactatcaagatgact----- 918
 6663 CTTCTAATCATGACAACGAGTACCTTCTCATCAAGATGACTGTGTAATAACATCTTGTGA 6722
 919 ----- 918
 6723 TCCACCCAGGAGGAGTCTTTGATAGTGTGCTGCTATCTCTTTGGCTCAACAGCTTTC 6782
 919 -----ggatatcctgtcatcaagaatagagaataaacctgaagactatat 964

6783 ATATGATTTTTCAGGTATATCTCTCATCAAGATAGAGAATAAACCTGAAGACTATAT 6842
 965 attgtatactaccgtggcgaaacagatcgttgggatggatatggtggtgcaagtgtgata 1024
 6843 ATTTGTATATACTACCGTGGCGAAACGATGCTTTGGATGGATATGTTGTCAGTTGTATA 6902
 1025 cagagaagtctgtattaccatagcattatccagaactcgaagaacagcagcaaaag 1084
 6903 CACGAGAAGTCTGTATTATCCCAATAGCATTTATACAGAACTCGAAAAAGCAGCAAAAG 6962
 1085 catagcagagacttcagcacattcattagaacgataacacatgtgttcctgaacctgc 1144
 6963 CATAGCAGAGACTTCAGCACATTCATTAGAACGGATAACACATGTTGTCCTGAACCTGC 7022
 1145 gctcgtggagaattgagaagacagtggaagaagtggaagaagtggaagaagtggaagaagtg 1204
 7023 GCTCGTGGAGAAATTTGAGAAGACAGTGAAGAAGTGAAGAAGTGAAGAAGTGAAGAAGT 7082
 1205 tgaagagatagaagaagagtagagaagaagtggaagaagtcggttagactgagatgac 1264
 7083 TGAAGAGATAGAAGAAGAGGTAGAGAAGGAGTGGAGAAGGTCGGTAGGACTGAGATGAC 7142
 1265 cttgttccagagattggtggaaggttttaataactgaagcagacgagaggaatttgc 1324
 7143 CTGTGTTCCAGAGATTGCTGAGGATTTAATGAACCTGAAGCAAGACGAGAGAAATTTCTGT 7202
 1325 gagagagtttaagtaagaagagatgaggttttggatgagatcaaaatggaagcagtgag 1384
 7203 GAGAGAGTTAAAGTAAAGAAAGAGATGAGTGGATTTTGGATGATGATCAAAATGGAAGCAAGTGA 7262
 1385 ggttgaataattgttggaaagctttgccaatcagaagagtcagtagaagaagaac 1444
 7263 GGTGAAAAATTTGTTGGAAAGCTTTGCCAATCAGAAAGGTCAGGTAGAAACAGAAC 7322
 1445 accattgtgtacaaactattatatactactgtgttcggttcataataaagtaatttt 1504
 7323 ACCATTGTTGTACAACTATATTATACATCTGTGTTGCTTCATATAAGTAATATTT 7382
 1505 tgtacacagtcacatccatccatacaaatggatgataaaaaa 1552
 7383 TGTACACAGTTCATCATCATTCCTCCATACAATTTGGATACAGAAAAACAA 7430

RESULT 3
ATU39452/c

LOCUS Arabidopsis thaliana syntaxin-related knolle mRNA, complete cds.
 DEFINITION Arabidopsis thaliana syntaxin-related knolle mRNA, complete cds.
 ACCESSION U39452
 VERSION U39452.1 GI:1184166

KEYWORDS thale cress.

SOURCE

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1 (bases 1 to 2596)

Lukowitz, W., Mayer, U. and Jurgens, G.

Cytokinesis in the Arabidopsis embryo involves the syntaxin-related

KNOLLE gene product

Cell 84 (1), 61-71 (1996)

JOURNAL

MEDLINE

96140640

REFERENCE

2 (bases 1 to 2884)

Lukowitz, W.

Direct Submission

Submitted (26-OCT-1995) Wolfgang Lukowitz, Lehrstuhl fuer

Entwicklungsgenetik, Universitaet Tuebingen, Spemannstrasse 37-39,

D-72076 Tuebingen, Germany

JOURNAL

FEATURES

1..2884

Location/Qualifiers

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

```
gene 520. .1452
/ gene="knolle"
CDS 520. .1452
/ gene="knolle"
/ note="syntxin-related"
/ codon_start=1
/ protein_id="AAC49163.1"
/ db_xref="GI:1184167"
/ translation="MNDLMTKSPMSYVDLKKAAKMDKAEAGDFLEWASTKADKMDEN
LSSLEAEYVKAEMGLSETLARIQYHEESKGVHKAESVKSIRKNSIELVSGLRK
AKSKLSKLEMDKANKETKRLSGTPYRSRTAVNGLRKLKKEVMFEQGLRQNMSE
YKEVERFYFTVGEHNDMEKLIIDNAGGEFLTRIQEHGKGVLETVVEIQDR
YDAKEIKSLEUHLQVFLDMVMVVEGQMDIEHHIVINASHYVADGANELKTAKS
HORSNRKWMCGIIGVLLIILIVIPITSSSS"
BASE COUNT 833 a 569 c 582 g 900 t
ORIGIN

Query Match 59.3%; Score 922; DB 14; Length 2884;
Best Local Similarity 91.4%; Pred. No. 5.5e-210;
Matches 1026; Conservative 0; Mismatches 5; Indels 91; Gaps 1;

QY 522 cagattaaatggggatctgtttgagacagtggtttgtgatgagttcaacagagtgact 581
Db 2845 CAGATTAAATGGGGATCTGTTTGAGAACACTGTTTGATGAGTTCACAGTGTGCT 2786

QY 582 gtgtcgagaaaaagtggttctgaataatctgatctcgagaaatttcctgccccagac 641
Db 2785 GTGTGAGCAAAAGTGTCTTCTGAGAAATCTGATCTCGGAGAAATTCCTGCCCCAGAC 2726

QY 642 ccttctgttctgacaaactcaactcgcgacttcaacgggaagtgggtacattaca 701
Db 2725 CCTTCTGTCTGTACAGAACTTCAACATCTCGGACTTTAAGCGGAAGTGTGATACATTACA 2666

QY 702 agtgccttgaatcccaacctttgtgcttcgactgcagctgcagtgatgagttccacacagaa 761
Db 2665 AGTGCCTTGAATCCCAACCTTTGATGCCCTTCGACTGCCAGTGCATGAGTCCACACAGAA 2606

QY 762 ggtgacacaaagctgttggaaacatctcttggagaataaagaccctagacagtgattc 821
Db 2605 GGTGACAAACAGCTTGTGGAAACATCTCTTGGAGAAATAAAGACCCCTAGACAGTGGATTC 2546

QY 822 ttactaggtcagccgtacaaaatctgtcgaagatctcaacacacacctgtgttctctac 881
Db 2545 TTTACTAGTCAAGCGGTACAAAATCTGTGCAAGATCTCAACCAACCTGGTGTCTCTAC 2486

QY 882 aatcatgacacagtgacttccatcaaatgact----- 918
Db 2485 AATCATGACACGAGTACTTCACTATCAAGATGACTGGTAAATAACATCTGTGTATCCACC 2426

QY 919 ----- 918
Db 2425 CACGAGGAGAGCTGTGNATAGTGTGCTATTTCTTTTGGCCCTAACAGCTTTTCATATGA 2366

QY 919 -----ggatatcctgtcatcaagatagagaataaaacctgaagcatatatattgt 970
Db 2365 TTTTTCGAGGTATATCTGTCTATCAAGATAGAGAAATAAACCTGAAACATATATTTGT 2306

QY 971 atactacgtggcgaaacagatgttgggatggatggatgggtggatgagttgtatcacagag 1030
Db 2305 ATACTACGTTGGCGAAACGATGCTTGGATGGATATGGTGGTGCAGTTGTATACACGAG 2246

QY 1031 aagttctgtattaccataagcattataccagaactcgaaagagcagcaaaagcatagg 1090
Db 2245 AAGTTCTGTATTACCCAAATAGCATATACAGAACTCGAAAGAGACGCAAAAGCATAGG 2186

QY 1091 cagagacttcagcattcattagaacggaataacacatgtggtcctgaacctgcgctcgt 1150
Db 2185 CAGAGACTTCAGCATTTCATTAGACGGATACACATGTGGTCCCTGAACTCGGCTCGT 2126

QY 1151 ggaagaaattgagaagacagtggaagagtggaagagataatcgtaaaagagttgaa 1210
Db 2125 GGAGAGAAATTGAGAAGACAGTGGAAAGAGTGAAGGATAATCGTAAAGAGGTTGAAGA 2066
```

```
QY 1211 qatagaagaagagtagagaagaagtgagaaagtcggttaggactgagatgacctgtgt 1270
Db 2065 GATAGAAGAAGAGGTAGAGAAGGAAGTGGAGAAGTGGTAGGACTGAGATGACCTTGT 2006

QY 1271 ccagagattgctgaagatttaatagaactgaagcagcagcagcagcagcagcagcagc 1330
Db 2005 CCAGAGATTGGCTGAAGGATTAACTGAACCTGAAGCAGCAGCAGCAGCAGCAGCAGC 1946

QY 1331 gtaagtaagaagagatgaggttttttgatgagatcaaaatggaagcagcagcagcagcagc 1390
Db 1945 GTTAAGTAAAGAAGAGATGAGTGTGGATGAGATCAAAATGGAAGCAAGTGAAGTGA 1886

QY 1391 aaaaattgttggaaagccttcccaatcaggaaggtcaggtgagaacacacacacacacatt 1450
Db 1885 AAAATTGTTTGGGAAGCTTTGCCAATCAGGAAGGTGAGTAGAAGCAAGCAACCAATT 1826

QY 1451 gttgacaaactatattacatactgttgcgttcacataataaataattttgtaca 1510
Db 1825 GTTGACAAACTATATTATACATACCTGTGTCGTTTCATATAAGTAATATTTTGTACA 1766

QY 1511 cagtcacatcattccatacaaatgtgataaaaaa 1552
Db 1765 CAGTCATCATCTTCATACATAAATGGATACAGAAACACAA 1724
```

RESULT 4

```
NTU34817
LOCUS NTU34817 1589 bp mRNA PLN 15-JUN-1998
DEFINITION Nicotiana tabacum violaxanthin de-epoxidase precursor (TVDE1),
mRNA, complete cds.
ACCESSION U34817
VERSION U34817.1 GI:1463122
KEYWORDS common tobacco.
SOURCE Nicotiana tabacum
ORGANISM Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
I; Solanales; Solanaceae; Nicotiana.
REFERENCE 1 (bases 1 to 1589)
AUTHORS Bugos, R.C., Hieber, A.D. and Yamamoto, H.Y.
TITLE Xanthophyll cycle enzymes are members of the lipocalin family, the
first identified from plants
J. Biol. Chem. 273 (25), 15321-15324 (1998)
98288256
REFERENCE 2 (bases 1 to 1589)
AUTHORS Bugos, R.C. and Yamamoto, H.Y.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-1995) Robert C. Bugos, Plant Molecular
Physiology, University of Hawaii, 3190 Maille Way, Honolulu, HI
96822, USA
```

FEATURES

```
Location/Qualifiers
1..1589
/organism="Nicotiana tabacum"
/strain="Xanthi"
/db_xref="taxon:4097"
/tissue_type="leaf"
1..1589
/gene="TVDE1"
42..1478
/gene="TVDE1"
/codon_start=1
/product="violaxanthin de-epoxidase precursor"
/protein_id="AAC50031.1"
/db_xref="GI:1463123"
/translation="MALAPHNSFLANHETIKYVVGSKI.PGHKREPSWGWEDYFGSIIVA
KICSSRLPRYFRKSPRICGLDSRGQLFSGHKNLSPAHSINQNVKNSGCKFPK
DVALWMEKQGFARFATLVAIFLVSADKADVALKTCTCLKECLELAKLSNPA
CAANVACLOTCNRPDETECOIKCGDLFENSVDDEFNECAVSRKCVPRKSDVGFV
PDPVLVKQKDMKDFSGKWEITRNLGTFDAFCQLHEFTENKLVGNLSWRITPD
GGFTTRSAQVQVODPKYPGILYNHDNEYLLODDWYILSKSVENSPEDYIFVYGR
NDWDGYGGSVLYTRSAVLPESILPELOTAQAKVGRDNFTIKTDNTCGPEPLVERL
EKVEEGERTIIKEVEIEEVEKVRKKEVTLFSKLFEGKFLQDRDENFLRELSKEE
```

MDVLDGLKMEATEVEKLFGRALPIRLR

transit_peptide 42..443

/gene="VDE1"

mat_peptide 444..1475

/gene="VDE1"

/product="violaxanthin de-epoxidase"

BASE COUNT 500 a 262 c 366 g 461 t

ORIGIN

Query Match 40.1%; Score 624.2; DB 14; Length 1589;

Best Local Similarity 74.3%; Pred. No. 8.2e-139;

Matches 820; Conservative 0; Mismatches 268; Indels 15; Gaps 2;

QY 364 ttgttccatctgcagatgcagttgatacacttaaaacttgatgcttattgaaggat 423

Db 424 TTGGTTCAAAAGCTGATGCGGTGATGCTCTCAAGACTTGTACTTCTGAAGAGT 483

QY 424 gcaggatagaactgcgaagtgcattgccaccctgctgtgcagccaatgctgcgtgc 483

Db 484 CGAGTTAGAGCTTGGCAAGTGCAATTCGAACCTGCTGATGTCAGCTTAATGTCCTGTC 543

QY 484 ttcagactgcataaocgtccagatgaacccgagtgccagattaaatgtgggactgt 543

Db 544 TCCAGACTTGCAACAATAGACCTGACGAACCGAATGTCAGATAAAATGTGGTCAATTGT 603

QY 544 ttgagaacagtggtgtgatgagttcaaacagtggtgctgtgcgagaaaaaagtgttc 603

Db 604 TTGAAAACAGTGTGCTGACAGAGTTCAATGAGTGTGCGAGTCTCCCGAAGAAATGTGTAC 663

QY 604 ctagaaaactgtatctcggaattctctgccagaccctctgtctgtctgtcagaact 663

Db 664 CTGTAATCTGATGTGGTGACTTCTCTGTACTGATCCAGTGTCTTGTGTCAGAGT 723

QY 664 tcaacatctcgactttaaagggaagtgtacattacaagtggcttgaatccaaacttg 723

Db 724 TTGACATGAAAGATTTTAGCGGGAATGTTCAATTACTCGCGGTTCGAATCCCACTTTTG 783

QY 724 atgctctgcagctgcagctgcagttccacacagagtgacacaaactgtgtggaa 783

Db 784 ATGCTTTTGTGTCGAATGTCATGAGTGTCCATACAGAA---GAAACAAACACTTGTGGGGA 840

QY 784 acatctcttgagataaaagccctagacagtgattcttactagctcagccgtacaaa 843

Db 841 ATTTATCTTGAGAAATACCTACACTGATGAGAGATTTTACTCGATCAGCGGTGCAAA 900

QY 844 aattcgtgaagatcttaaccaactggtgtctctcaatcatgacaacaggtaccttc 903

Db 901 AATTGTCGAAGATCCAAAGTATCCGGGATACCTTCAATCATGATGATGATCTTC 960

QY 904 actataagatgactggtatatactctgcatacaagatagagataaaacctgaagactata 963

Db 961 TCTACCAAGATGACTGTTATATTTTGTATCCAAAGTAGAAAAATAGTCCAGAGATTACA 1020

QY 964 tattgtatactacgtggcgcaacatgctgtggatgatatggtggcagttgat 1023

Db 1021 TATTGTGTACTAATAGGGCGAAGTATGATGCGGATGGATGATGTTGTTGTTCTACTTT 1080

QY 1024 acagagaagttctgtattaccatagcattatcaccagaactgcgaagcaagcaaaaa 1083

Db 1081 ACACAAGAGTGCAGTTTGTGCTGAAAGCATTATACCGAGTTTCAAAACCGCAGCTCAA 1140

QY 1084 gcattagcagagacttcagacatactattagaacgagataacacatgtggtcctgaacctg 1143

Db 1141 AAGTTGGCGGTGATTTCACACATATCAATAAAACAGACAATACATGATGTCGCTGAACCTC 1200

QY 1144 cgtcgtgagagaattgagaagcagtggaagagtggaagataatcgtataaacagg 1203

Db 1201 CCTTTGTGAGAGGTTGGAGGAAGAGTGGGAAGAGGAGGACCATCATATAAAGAG 1260

QY 1204 ttgaagagatagaagaagtagagaaggaagtggagaaggtcgtgaggactgagatga 1263

Db 1261 TTGAGGAGATAGAAGAAAGTAGAGAGGTGAGAGATAAAGAG-----TCA 1308

QY 1264 cctgttcagagattgctgaagatttaatagaactgaagcaacagagagaatttcg 1323
Db 1309 CCTATTTCAGTAAACTGTTTGAAGGTTTAAAGAGCTCCAACGAGATGAAGAGACTTCT 1368
QY 1324 tgagagagtaagtaagaagagagtgagtttttgatgagatcaaaaatgggaagcaagt 1383
Db 1369 TAAGAGAGCTGAGCAAGAAGAATAATGGATGTTTGGATGACTTAAATGGAAGCAACTG 1428
QY 1384 aggttgaaaaattgttgggaagcttttcccaatcagggaaggtcaggttagaacaagaac 1443
Db 1429 AGGTGAAAAAATCTTTTGGCGCTGCTTTACCAATAAGGAATAATTAAGGTAAATTTTAA 1488
QY 1444 caccattgtgtacaaactatat 1466
Db 1489 AACTATCAACATATATACATACAT 1511

RESULT 5

LSU31462

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

transit_peptide

CDS

gene

mat_peptide

BASE COUNT

ORIGIN

PLN 22-JUL-1996

mRNA

1760 bp

Lactuca sativa violaxanthin de-epoxidase (VDE1) precursor, mRNA,

complete cds.

U31462

U31462.1

GI:1438874

Lactuca sativa.

Lactuca sativa

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.

Bugs, R.C. and Yamamoto, H.Y.

Molecular cloning of violaxanthin de-epoxidase from romaine lettuce and expression in Escherichia coli

Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6320-6325 (1996)

96270536

2 (bases 1 to 1760)

Bugs, R.C.

Direct Submission

Submitted (12-JUL-1995) Plant Molecular Physiology, University of Hawaii, 3190 Maile Way, Honolulu, HI 96822, USA

Location/Qualifiers

1..1760

/organism="Lactuca sativa"

/strain="Romaine"

/db_xref="taxon:4236"

235..609

/gene="VDE1"

235..1656

/codon_start=1

/product="violaxanthin de-epoxidase precursor"

/protein_id="AAC49373.1"

/translation="MALSLHTVFLCKEALNLYABSPCNERFHRSGQPPTNIMMKIR

SNGGFNSFRFTSTKTSFSDSSKCKSQICSDTSTFEIQRDLKRGMTLLLEKQ

WQFQALAILVCTEVIIVRVDAVDALTKACLLKREKRIELAKIANPSKANVACLQ

TCNNRPDETECIKCGDLFENSVDQFNCAVSRKKCPVKSDVGEFVPPDRNAVON

FNKDFSGKWTITGLNPTDFADQDLHEFHMENDKLVNLTWRITKLDGFFTRSAV

OTFYODPLPGALYNHDFEFLHYODDWILSSOIKENKPDYIFVYGRNDADWDYGG

SVIYRSTPLPESIIPNLQKAQSVGRDNFNTTDSNGSGPEPPLVERLEKTAEGEK

LLIKGAVEIEEVEKEVERKRDTEMTLFRLEGLFKLQOQDENFVRELSKEEKLIN

ELQMEATEVEKLFGRALPIRLR"

235..1656

/gene="VDE1"

610..1653

/gene="VDE1"

/product="violaxanthin de-epoxidase"

557 a 319 c 387 g 497 t

Query Match	39.3%;	Score 611.8;	DB 14;	Length 1760;
Best Local Similarity	72.5%;	Pred. No. 7.6e-136;		
Matches 806;	Conservative 0;	Mismatches 302;	Indels 3;	Gaps 36;
Qy	324	ttgctaaactcgtggggtttttagcttgcgcgtctccctattgtttccatctgcagatgca	383	
Db	550	TTcATAcAAATTTGCTATGCTATTGGTTTGcAcATTTGTTATGCTTCCcAGAGTTGATGCC	609	
Qy	384	gttgatgcacttaaaactttgctatgctttattgaagggtatcgagatagaactgcgaag	443	
Db	610	GTTGATGCTCTTAAAcCTTGCTGTTTACTCAAGAATGCAGGATTGAGCTTGCAAAA	669	
Qy	444	tgaattgccaacctccctgtgcagccaatgctgcggtgccttcagacctgcaataaacgt	503	
Db	670	TGtATAGCAAAACCCATCTTGTCGGGCAACGTGTGCTGTCTAcAGACTTGCACAAATCGT	729	
Qy	504	ccagatgaaccaggatgccagatataaattggggatctgtttggaaacagttgttgtgat	563	
Db	730	CTTGACGAGACCGAATGTcAGATAAAATGTGTGACTTGTTCGAAAcAGTGTGGTGAC	789	
Qy	564	gagttcaacgagttgctgtgtgcgaaaaaaagtgtgttctctagaaaaatctgactcggg	623	
Db	790	CAATTTAAACGAGTGTGGGTTTCCCGAAAGAAATGTGTGCCCGGAAATCGGATGTGGGT	849	
Qy	624	gaatttctgtccccagaccctctgttctgttacgagaacttcaacatctcgacttttaac	683	
Db	850	GAATTCCTCGGTTCCGGATCGTAATGCAGTGGTTCAAATTTTAACATGAAGACTTTAGT	909	
Qy	684	gggaagtgttacattacaagtggcttgaatacacaaccttgcgttcctgcagctgcagctg	743	
Db	910	GGGAAGTGGTATTATAcAAGTGGTTTAAATCTCTACATTTGATGTGATTTGTCAACTT	969	
Qy	744	catgagttccacacagaaggtgacacaagctgttggaacaactctcttggagaataaag	803	
Db	970	CATGAGTTTCATATGGAATG--ATAAACTGTGGGACTTAAcATGCGCGATAAAA	1026	
Qy	804	acctagacagttgatatcttactagggtcagcgcgtacaaaaatbtgtgcaagatcctaac	863	
Db	1027	ACTTTGGATGGTGGTTCTTTACTCGATCTGCTGTGCAAAcATTGTTTCAAGATCCAGAT	1086	
Qy	864	caacctgtgttctcacaatcatgacaacgagtcaccttcactacaagatgactggat	923	
Db	1087	CTTCTCGGACACTTTATTAATCATGACAATGASTTCTTCTACTACCAAGATGACTGGTAC	1146	
Qy	924	atcctgtcatcaagatagagataaaacctgaagactatatattgtatactaccgtggg	983	
Db	1147	ATATTATTTCTCCAAATCGAAAcAAACCCGATGATTACATATTCTGTATCTACCGAGGT	1206	
Qy	984	cgaacgagcttggatgatatgtgtgtgcagttgtatcacagagaagttctgtatta	1043	
Db	1207	CGAAACGACGcATGGATGGATACGGTGGGTCCGTGTACTACCCGAAGCCCGACACT	1266	
Qy	1044	cccaatagcattataccagaactcgaaaaagcagcaaaaagcataggcagagacttcagc	1103	
Db	1267	CCCGAATCGATCATCTCCAAACCTACAAAAAGCAGCCAATCCGTGGGTGAGACTTTAAC	1326	
Qy	1104	acattcattagaacgataaacacatggtgctgcgaacctgcgctgtggagagaattgag	1163	
Db	1327	AATTTTTATAAACCCGACAAATAGTTTGTGGGCTGAGCTCCATTTGGTGGAAAGCCTTGAG	1386	
Qy	1164	aagacgtggagaaggtgaaaggataaactgtaaaagsggttgaagagatagagaagag	1223	
Db	1387	AAAACAGCGGAAGGCGAGAAGTTGTGTATAAAGAAGCTGTAGAGATAGAGAAGAG	1446	
Qy	1224	gtagaagaaggatggagaaggttcggtgaggaactgagatgacctgttccagagattggct	1283	
Db	1447	GTTTCAAAAGAGGTTGAGAAGGTTAGATATCTGATCTGAGTGTGCTTTGTTCAGAGGTTGCTT	1506	
.Qy	1284	gaaggatttaatgaactgaagcaacgacgaggaatttcgtgcaagagtttaagttaagaa	1343	
Db	1507	GAGGGGTTTAAGAGGTTTGCACACAGATGAAGAGAATTTTGTGAGGGAGTTGAGTAAAGAA	1566	
Qy	1344	gagatggagtttttggatgagatcaaaatggaagcaagtcaggttgaaattgttttggg	1403	

D	b	1567	GAGAGGAAATTCCTGAATGAACCTTCAAATGAACGGACTGAAGTTGAAAAGCTTTTGGG	1626
Q	y	1404	aaagttgccaatcagaaggtcaggtaga	1434
D	b	1627	C GCGGTACCATTAGGAACTTAGATAA	1657
<hr/>				
RESULT 6				
SOL250433				
LOCUS	SOL250433	1886 bp	mRNA	PLN 21-OCT-1999
DEFINITION	Spinacia oleracea mRNA for violaxanthin de-epoxidase (svdel gene).			
ACCESSION	AJ250433			
VERSION	AJ250433.1	GI:6103242		
KEYWORDS	svdel gene; violaxanthin de-epoxidase.			
SOURCE	Spinach.			
ORGANISM	Spinacia oleracea			
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales; Spinacia.				
REFERENCE				
AUTHORS	Emanuelsson,A.K., Eskling,M. and Akerlund,H.E.			
TITLE	Cloning and sequencing of Spinacia olercea violaxanthin de-epoxidase			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 1886)			
AUTHORS	Emanuelsson,A.K.			
TITLE	Direct Submission			
JOURNAL	Submitted (20-OCT-1999) Emanuelsson A.K., Dept. of Plant Biochemistry, Lund University, Chemistry Center, POB 117, 221 00 Lund, SWEDEN			
<hr/>				
FEATURES				
source	Location/Qualifiers			
	1..1686			
	/organism="Spinacia oleracea"			
	/db_xref="taxon:3562"			
	/country="Sweden"			
	/tissue_type="leaf"			
sig_peptide	1..372			
	/gene="svdel"			
CDS	1..1419			
	/gene="svdel"			
	/function="part of the xanthophyll cycle"			
	/codon_start=1			
	/product="violaxanthin de-epoxidase"			
	/protein_id="CAB59211.1"			
	/db_xref="GI:6103243"			
	/translation="MALVARISCVYEYIAGICNNVSHRNFKKWQKNPFLFODDARNRIFNRDKLSCTFFKASPKLOHSPKSGLLISCGEVNSSKVSNNAVIPKKWNLLKLVKEVTIVACTCFWFGSSAOVDALKTCTLLKEGRLELAKNCIANPSCAANVACLOT CNRPDETECOIQGLDFANKYVDEFECAVSRRCKVPOKSDVGCFEPDPSPVLVSFNMADFNKGWPISSLNGTDFAFCQLHEFDLGKLGNLSWRIKTPDGGFPTRTAVQ KFAQDSQPQMLYNHYNAYLHQDDWYILSSKIENPDPPYVFVYYRGRNDAMDGYGGA FLYTRSATPEVNIPVAQAQGVGDNFNKIRTDNTCGPEPPLVERLEKTVBEGERIT IIEKVEGLEIEIGDLKVKGTBMTLFQRLLGEFQELQKDDEYFLKLNKERELLEDD LKMDAEVEKFLGALPIRKLR"			
gene	1..1419			
	/gene="svdel"			
mat_peptide	373..1416			
	/gene="svdel"			
	/product="violaxanthin de-epoxidase"			
BASE COUNT	549 a 296 c 375 g 466 t			
ORIGIN				
<hr/>				
Query Match 38.7%; Score 601.4; DB 15; Length 1686;				
Best Local Similarity 71.8%; Pred. No. 2.3e-133;				
Matches 801; Conservative 0; Mismatches 311; Indels 3; Gaps 1;				
Q	y	318	ccgcttgttgataaacctcggtgttttagcttcgcgttccttatgttcacatgca	377
D	b	307	CTGAAGGTGTGTAAGTGACTGCTATAGTCCGCTGCATCTTTTGTGTCATGCTTCGGCA	366
Q	y	378	gatcgactgtgatcacttaaaaacctgtgcattgttcattgaaggatcgagcatagaactc	437

	Query Match	38.7%	Score 601.4;	DB 15;	Length 1686;
	Best Local Similarity	71.9%;	Pred. No. 2.3e-133;		
	Matches 801; Conservative	0;	Mismatches 311;	Indels	Gaps 1;
Qy	318 ccgcgtgctgataaacctcggtgggtttttagcttgccgttccttatgttccatcgcga	377			
Dd	307 CTGAAGGTGTTTGTAAGTGCACGTCTATAGTTGCGGTACATTTTTTTCATGCTCCTCGGCA	366			
Qy	378 gatcagcttgatgcacttaaaacctgtgcagtcttatgaagggatcgaggatagaactc	437			

Db	367	CAAGCTGTTGATGCTCTTAAAGCTTGTACCTGTTTCTGAAGCAATGCGAATTTGAACATC	426
Qy	438	qcaaatgcatgccaacccctgctgagccaatgctgcgcctcagaccctgcaat	497
Db	427	GCCAAAGTCGCAATCAATCTTGTGAGCAATGTTGCTTGCCTGCAGACCTGCAAT	486
Qy	498	aaccgtccagatgaaccgagtgccagattaaatgtgggagctgtgttgagacagtggt	557
Db	487	AATAGACCTGATGAACCGAATGCCAAATCAAAATGTGGGACTGTTGCGCAACAAGTT	546
Qy	558	gttgatgagtcacacgagtgctgtgtgcgagaaaaaagtgtgttcttagaaaaatcgtat	617
Db	547	GTTGACAGCTCAACGAGTGCAGTCTCTCGAAAGAAATGCGTGCTCAGAAAGTCTGAT	606
Qy	618	ctcgagaaattctctgccagaccctctgttctgtacagaaactcaacatctcggac	677
Db	607	GTTGGAGAATTCCTGTCCGATCCTAGTGTGCTGCTTAAAGAGTTTCAACATGGCAGAT	666
Qy	678	ttaacgggaagtgtgtacatacaagtgtgtgaatccaacaccttgatgccttcgactgc	737
Db	667	TTCACGGGAAGTGGTTTATAAGTAGTGTCTTAACCCCTACATTCGACGCTTTTGATTC	726
Qy	738	cagctgcatagttccacacagagtgacacaaactgttggaacatctcttgaga	797
Db	727	CAGTTACATGAGTTCATTTGGAGATG---GAAACCTTGTGGAACTTGTCTTGCGGA	783
Qy	798	ataagaccttagacagtgattcttactaggtcagccgtacaaaaattcgtgcaagt	857
Db	784	ATAAAACACAGATGCTGTTTTTTCACAGCAGCTCTGTACAGAAATTTGCGCAAGAC	843
Qy	858	cctaaccacacctgtgtctctacacatcatgacaaacagtagtaccttcaactcaagatgac	917
Db	844	CCCTCTCAACCTGGAATGCTGTATAATCATGACAATGCATATCTTCACTATCAAGATGAT	903
Qy	918	tgttatctctgtcatcaagatagagaataaacctgaagcatatatttgtatactac	977
Db	904	TGGTACATCTCTTCTTAANTTGAANTTCAACCAACGACGACTACGATTTTGTATATAC	963
Qy	978	cgtggggcaaacgtagtgggtgatgtgtgtgtgagttgtgtgtatatacacagaaagttct	1037
Db	964	CGAGCGAGGAATGATGCTGGGATGGCTATGTTGGTGGCGCATTTCTTTACACAAGAAGTGA	1023
Qy	1038	gtattaccaatagcattatcacgaactcgaaaaaacagcaaaacagcatagcgagac	1097
Db	1024	ACTGTACCTGAAAATATGCTGCCCTGAACCTTAACAGAGCAGCTCAAAAGTGTAGGAAAAGAC	1083
Qy	1098	ttcagacattcattagaacgataacacatgtgtgtcctgaacctgcgctcgtggagaga	1157
Db	1084	TTCATAAATTCACGACGCAATACCTGTGGGCCCGAGCTCCACTGGTTGAGAGG	1143
Qy	1158	attgagaagacagtggaagagtgaaaggataatcgtataaagaggttgaagagatagaa	1217
Db	1144	CTGGAGAAGACGGTGGAGAGAGGAAAGAACTATCATTAAGGAAGTAGAACAATTAAGAA	1203
Qy	1218	gaagaggttagaaggaagtgaaggtcgttaggactgaagtagactgttccacaga	1277
Db	1204	GGAGAGATAGAGGAGACCTGGAAAGGTTGGGAAAACCTGAGATGACATGTTTTCAGAGG	1263
Qy	1278	ttggctgaaggttaataatgaactgaacagacgagagagaatttcgtgagagagtttaagt	1337
Db	1264	CTACTAGAGGTTTCAAGACCTCAAAAAGATGAAGATACTTCTCTCAAGAATTAAC	1323
Qy	1338	aaagaagagatggagtttttggatgagatcaaaatggaagcaagtgaaggttgaataattg	1397
Db	1324	RAGGAAGACGAGAAATTAAGTGAAGACTTAAAGATGACGCTGGCGAGGTGGAAAAACTA	1383
Qy	1398	tttggaaagcttttccaatcaggaagtcaggta	1432
Db	1384	TTTGGGCTAGCATTAACCTATAAGAAAGCTTCGGTA	1418

Db	238	CGCACCCCGACTCTGGCTCTTCCACGAGACGACCATACAGCGGTTGTGTCAGGACCCA	297	REFERENCE	Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
QY	861	aaccaactggtgtctctactaatcatgaacagagtcaccttactactatcaagatgactgg	920	AUTHORS	Ecker, J.R.
Db	298	GCACAAACCGCGATCCTCTATACCACTACACACGAGTTCCTCGCATATCAAGATGACTGG	357	JOURNAL	Submitted (20-MAY-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
QY	921	tatatctctcatcaaaagatagagaataaaactgaagactatatattgttatactacgt	980	REFERENCE	Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
Db	358	TACATATCTCATCAAGTAGAGACAAAGAGATGACTACATTTTCGTATACCGC	417	AUTHORS	Ecker, J.R.
QY	981	ggcgaaacgatctgtggatgatattgtgtgcagttgtatcacagagaagttctgtga	1040	JOURNAL	Submitted (30-OCT-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
Db	418	GCAGAAATGATGCGTGGGTGCTATGCTGTCTGTCTGTACACCAAGCAAGATT	477	REFERENCE	Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
QY	1041	ttaccaatagattatcaccagaactcgaataaagcagcaaaacatagaggagagacttc	1100	AUTHORS	Ecker, J.R.
Db	478	GTGCTGTATCATATAGTACCTGAGCTAGAAAGGCGCGCAAAAGCGTACGTCGCGACTTC	537	JOURNAL	Submitted (25-NOV-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
QY	1101	agcacattcattagaacggaataacacatgtgtctctgaacctgcgtcgtagagaatt	1160	REFERENCE	Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
Db	538	TCCAGCTTCATCAGGACCGACCAACACCTGCGTCCCGAGCCACCTCTCGTGAGAGATA	597	AUTHORS	Chao, Q., Brooks, S., Buehler, E., Johnson-Hopson, C., Khan, S., Kim, C., Shinn, P., Altai, H., Bei, B., Chin, C., Chlou, J., Choi, E., Conn, L., Conway, A., Gonzalez, A., Hansen, N., Howling, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R., Federspiel, N., Theologis, A. and Ecker, J.
QY	1161	gagaagacagtggaagaagtgaaagataatctaaagaaggttgagagataagaagaa	1220	TITLE	Direct Submission
Db	598	GAGAAAGCCGTGCGAAGGAGAGAACCATCATCAGGGAAGTGCAGGAGATAGAGGGC	657	JOURNAL	Submitted (03-JAN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, PA 19104-6018, USA
QY	1221	gaggtagagaagagtgaggaagtcggttaggactgagatgacctgtgtccagagattg	1280	REFERENCE	Hamilton Walk, Philadelphia, PA 19104-6018, USA
Db	658	GAGATCGAGGAGAGGTGAAGGACTGGAGGAGGAGGAGTCACTATGTTCAAGAGTTG	717	AUTHORS	Chao, Q., Brooks, S., Buehler, E., Johnson-Hopson, C., Khan, S., Kim, C., Shinn, P., Altai, H., Bei, B., Chin, C., Chlou, J., Choi, E., Conn, L., Conway, A., Gonzalez, A., Hansen, N., Howling, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R., Federspiel, N., Theologis, A. and Ecker, J.
QY	1281	gctgaagatttaactgaactgaacgaagcagaggaatttcgtgagagattgaagtaaa	1340	TITLE	Direct Submission
Db	718	ACAGTGCCCTCATGGAGTGAACAGGACTTGATGAACTTCTTCAGGCGCTGAGCAAG	777	JOURNAL	Submitted (11-OCT-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, PA 19104-6018, USA
QY	1341	gaagagatggatttttgatgatcaaatgaacgaagcaggaatttcgtgagagattgaagtaaa	1400	COMMENT	On Nov 25, 1999 this sequence version replaced gi:6143825.
Db	778	GAGGATGGAGCTTTGGATCATGATGACATGACATGGAACCACTGAGTTGAGAGGCTTC	837	FEATURES	Location/Qualifiers
QY	1401	gggaagctttgccaatcagaaggttcaggtagaacaag	1440	source	1. 89479
Db	838	AGCGCGCCCTGCCAATAAGGAATAAGGTAGTCTTAG	877	CDS	/organism="Arabidopsis thaliana" /db_xref="taxon:3702" /chromosome="1" /clone="T27G7" join(<1..126,240..382,525..723,861..1172) /note="unknown protein gb AAD21437.1; similar to EST gb N96004"
RESULT	8				
AC006932					
LOCUS	AC006932	89479 bp	DNA	PLN	11-OCT-2000
DEFINITION	Genomic sequence for Arabidopsis thaliana BAC T27G7 from chromosome I, complete sequence.				
ACCESSION	AC006932				
VERSION	AC006932.8	GI:6468048			
KEYWORDS	HTG.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
AUTHORS	1 (bases 1 to 89479)				
	Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C., Khan, S., Kim, C., Altai, H., Bei, B., Chin, C., Chlou, J., Choi, E., Conn, L., Conway, A., Gonzalez, A., Hansen, N., Howling, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Federspiel, N., Theologis, A. and Ecker, J.				
TITLE	Genomic sequence for Arabidopsis thaliana BAC T27G7 from chromosome I				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 89479)				
AUTHORS	Ecker, J.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-MAR-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and				


```
KKAYVOOLESSRIKLSOLEQELQARSOGLFMGCGPPGNITSGAAIFDMEXCRWLE
DDNRHMSIRGLQAHSLDDNRLRLIVDGYIAHDEIFRLKAVAKADVFHLITGTWMS
PAERCFIWAAGFRSDLIKILVSOQMLDLEQLMGIIYSLQHSQOAEALSQGLEQIQ
OSLIDTAAVFDVDMQOMAVAGLISNLECFIRQADNLRQOVTVOLRLITVRAQAR
FLVIGETVGLRLSLMSLRSDSVSEQTRAAALAEALRLISKQDPSRLITADAPA
IPYLAETLYSSHSOENAAATLNLSTREPLMSRGLDLSHALRHDDTTSPA
AVOSSAATYSLIIAEYRPIIGSKRDIIFSLHIIRYDPSHPRSIKDSLKALFAIA
LYPNRSTMSIIGAEPALFSLIVDSRCGVEDATAYMAQVAGCEDSEDEGMRVRSGV
VLADLDCTSSURIKENSVALNLCARCGAAARSEVAARVAGCEDEGMRVRSGV
AENSLGKRKAVALLKLVVSGNGGDSRDYFENENPNRSR"
complement(join(6389..6646,7024..7106,7608..7702,
8179..8287,8565..8654,8797..8890,9113..9184))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="T27G7.3"
/protein_id="AAF22905.1"
/db_xref="GI:6664323"
/translation="MNRDRPLHLINCOVLASPLTAIKLIEKGGEKERVMNHKGFY
KYRNSKERTIKKPLTRKLPFLSILVLKHOKNITKEMANHRMSATNHNH
HLPTSLHGLNNHPSSGFINQDSSSDFGELEEAIVLQGVKYRNEEAAPPLLGGG
GATLEMPMSPIRTHQTLPTESKSGESDGSANFSQKESQSPESMSKHHLM
LQPHNNMANSSTSGLPSTSRTLAPPKPSDEKVLIIIIIIIIIF"
join(11373..11439,11990..12266,12331..12619,12691..12921,
13105..13503)
/note="putative rac GTPase activatin protein gb|AAD15596;
similar to ESTs gb|T20436"
/codon_start=1
/evidence=not_experimental
/product="T27G7.4"
/protein_id="AAF22885.1"
/db_xref="GI:6664303"
/translation="MGHILWIFASGPYFGLLDCNSGFARGETIERLKKLKSLEWFG
GSTSVGRVLELAEFAVEDNDPAMDIGGPTNIRHVAHVTFDFDGLGLPSE
FEPVQKAPARHIIILVFGSATVFGDSTESQMSYDSRGNCVPIVLLLSRLY
DOGLQAEVPRITGENSEEFVREQLNKGIIDPGIDVHCLAGLIKVLVIAFREL
RGVLDLPSEVOMCESDEDFKVRLLPQTEASLLNAINLMADVIOFEHVKMSR
NLALVFAPNSQADPLTALMAVQVVKLLKSLTEKTVREREASSVVDRCSEKED
GEKEDNEEREDEEEDDEDEEEDGDDVYIIKEEASEIIVKVADEHKSGSI
KSEFEGSATDSKGDNGVQPPICSNP"
complement(join(14188..14733,14829..14985,15122..15253,
15357..15945,16043..16083,16294..16439))
/note="putative multispinning membrane protein
gb|AAD20090"
/codon_start=1
/evidence=not_experimental
/product="T27G7.5"
/protein_id="AAF22904.1"
/db_xref="GI:6664322"
/translation="MAOFLTLVQLVLLATFWIGIGSSSNHYNAGDHVPLVYNKVP
LHNPSEYVYDLPFCRRGPVIEKQETLGVNLGDRMLSSLYKLKFPREDKTHFVLCK
RLTSDIARFDIIAQDIYFOMYDDILPLMGVKGVEGDFGQGERHRTYIIFSHLF
NVLNADKIEINSFSDSYMVIDISEIDVQFTYSVSNLTSERSETRMNYSRAS
FHPISQIHFFSNTSVVYLLGLISFLFMRLKNEKLSYISIGDEERKEAGWLV
HSDVFRCPNLSILGTLGTQLLIIIALFALFTFLYPYNRGLMLTSLVIMTLT
TSIVAGTYSFHSQFNGKQKRSVLGILYVPVFFIILSVLNTVAITVGAAPLP
GTIVIIILFTLLNPLMLGLVGNFGLLEFQPPSAVXRNPREIPQWYRKLIQ
VFLGFGFPFSAVLEWQLIASLGMFKFIITSPGIMLFTFVILFSSSVGIILYIQL
SGDEHWMRVRYFLYIFAFPSIQTCFDP"
join(17065..17069,17482..17701,17921..18116,18380..18528)
/note="putative ribosomal protein L10 gi|3860277; similar
to ESTs gb|T41543, gb|R86781, gb|AA721952, emb|Z34253,
gb|N64989, gb|AA041179, emb|Z26567"
/codon_start=1
/evidence=not_experimental
/product="T27G7.6"
/protein_id="AAF22886.1"
/db_xref="GI:6664304"
/translation="MSLKQSEAVREAITTTITGSKAKRNFVETIELQIGLKNVDPOK
DKRFSGVSLPHIPRPMKICMUGDAQHVEAEKMGLENMDVSLKKNKLVKLL
AKKHAFIASSSVIKQIPRLGPNLAKGKFWLQILFHFNLNLSMEEKQIFQNVQMSYN
FLVSLKKNQNVRLYIKSTMGPQRI"
join(19189..19347,19428..19605,19741..19885,20087..20143,
```

```
20295..20409,20600..21079,21145..21201)
/note="similar to yeast dcpl gi|3169719; similar to ESTs
gb|H37036, emb|Z27040, gb|R90173, gb|AI999440.1"
/codon_start=1
/evidence=not_experimental
/product="T27G7.7"
/protein_id="AAF22887.1"
/db_xref="GI:6664305"
/translation="MSONGKIIPNLQDNSTRLLNLTVLQRIIDPYIEILITAAHVTY
EFNIELSQSRKDVESLFWKRLLSIFNYVLIIFNLKSDWILFFVFSRSTOPRF
QFTVMNRNLDNLVLLGDFEYVQGVYLLRNASQEVNGIWFYNKRCEBVALFN
RIILSAYKVNOKPKASSKSEFEELAKPTMAVMOGLPESTARDAPDPAVAFES
SYMNLGNTASGSASGPYQSAIPHQPHQPHQPIAPPVAAAAPPQIQSPPPQSSPL
MTLFDNNPVISSNHIHDLVTPSEFGPPRMAQPHLIPGVSMPTAPPENPNNAHQ
QRSYGTPIVLPQPPPTPPPSLAPAPTGPVISRDKVEALLSLQDEDEFIDKTRTLQ
ALQO"
complement(join(24182..26122,26212..26300,26373..26501,
26545..26839))
/note="similar to inerleukin-1 beta convertase homolog A
precursor sp|P55865"
/codon_start=1
/evidence=not_experimental
/product="T27G7.8"
/protein_id="AAF22903.1"
/db_xref="GI:6664321"
/translation="MLPMEHPLPLSVLPNPAITLPGIALGFDIGNFVLDLILLRAST
LTSNLNHDSCDLNDRLLHRLTLTKHAVSWISTLSAKVSLDLRLNLSLCLHIEF
Query Match 21.2%; Score 330.4; DB 12; Length 89479;
Best Local Similarity 83.5%; Pred. No. 2.5e-68;
Matches 421; Conservative 0; Mismatches 6; Indels 77; Gaps 1;
QY 2 cagcgctccggtggtggtgggaagattagatagtagtgagaatggcagtagctacaca 61
Db 88259 CTCAGGTATTGCTGGTGGGAAGATTAGATAGTGTGAAGAATGCGACTAGTACACA 88318
QY 62 ttgtttcaacttcactgcatgacccgattatcgattttttcgaagtagatgattgg 121
Db 88319 TTGTTTCACCTTCACCTTGTTCATGACCGTATTCGATTTCCTCAAGTGATGATTGG 88378
QY 122 taggcttggtcattacaagaagagatcaatggcactttcttctcaagattttacctoc 181
Db 88379 TAGGCTTGGCATTACAGAAGAAGAGGATCAATGGCATTCTTCTGCTCAAGATTTTACCTCC 88438
QY 182 aatccaaagtgtgtagctcagacaaactggtgggagatcctcagctccttattctgatt 241
Db 88439 AATCCAAAGTGTGATCTCAGAACAACTGTTGGGAGATCCTCAGCTCTTATCTGCATT 88498
QY 242 caggtcaggattctcta----- 258
Db 88499 CAGGTCAAGGATTCCTTAAGTAATCAATCACTAGTCTTGTGTTCATACAAAATTTG 88558
QY 259 -----aggggatatttgacatttgccatta 284
Db 88559 GTATTTCGATGATCTCATTGCTTTTGTGATCATGTCAGGGGATATTTGACATTGTGCCATTA 88618
QY 285 ccataaagaatagctggaagagctgaccgctccgctgtgtgctaaaaactcgtgggtgtt 344
Db 88619 CCATCAAGAAGTACGCTGGAAGAGCTGACCGCTCGCTGTGTGCTAAAACCTCGTGGGTGT 88678
QY 345 ttagcttgccgttccttattgttcactatgcagatgcagatgcagatgcacitaaacttgt 404
Db 88679 TTAGCTTGCCGCTTCCCTATTGTTCATCTGCAGATGCAGTGTGATGCACCTTAAACTTGT 88738
QY 405 gcatgcttattgaaggatgcagg 428
Db 88739 GCATGCTTATTGAAGGATGCAGS 88762
RESULT 9
166494 7218 bp DNA PAT 28-DEC-1997
LOCUS
DEFINITION Sequence 14 from patent US 5670367.
```

```

ACCESSION   I66494
VERSION     166494.1  GI:2724471
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 7218)
AUTHORS     Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE       Recombinant fowlpox virus
JOURNAL     Patent: US 5670367-A 14 23-SEP-1997;
FEATURES    Location/Qualifiers
             source
             1..7218
             /organism="unknown"
BASE COUNT  1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match          5.9%; Score 92.4; DB 10; Length 7218;
Best Local Similarity 3.1%; Pred. No. 1.4e-11;
Matches 12; Conservative 255; Mismatches 121; Indels 0; Gaps 0;

Qy 1056 ataccagaactcgaaagcagcaaaagcatagggagagacttcagcacattcattaga 1115
      ||| ||| ||| | | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1451 ATACAAGAAATTGCTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1392

Qy 1116 acgatacacatgtgtctcgaacctgcgtgagagaaattgagagacagtgga 1175
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1391 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1332

Qy 1176 gaagtgaaaggataatcgtaaaagaggttgaagagatagaagaagaggtagagaaggaa 1235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1331 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1272

Qy 1236 gtggagaggtcggtagactgagatgacattgttcagagattggtcgaagattaat 1295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1271 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1212

Qy 1296 gaactgaagcaagcagggagaaatttgtgagagagttgaagtaagaagagatggagttt 1355
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1211 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1152

Qy 1356 ttgtagatgacaaatggagcaagtgaggttgaaaaattgttggaaaagatttgcca 1415
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1151 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1092

Qy 1416 atcaggaggtcaggttagaacaagaac 1443
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1091 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1064

RESULT 10
HSAJ9612
LOCUS       HSAJ9612 147243 bp DNA HTG 31-MAY-2000
DEFINITION Homo sapiens chromosome 17 clone PAC RPCI-1 77H15 map 17p11.2, ***
SEQUENCING IN PROGRESS ***, 7 unordered pieces.
ACCESSION  AJ009612
VERSION     AJ009612.4  GI:9211528
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 147243)
AUTHORS     Radelof, U., Hennig, S., Ramser, J., Francis, F., Steffens, C.,
             Klein, M., Seranuk, P., Poustka, A., Reinhardt, R. and Lehrach, H.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 147243)
AUTHORS     MPIMG.
TITLE       Direct Submission
JOURNAL     Submitted (17-JUL-1998) MPIMG, Abt. Lehrach, Max Planck Institut
             fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195, Germany
             On Jul 15, 2000 this sequence version replaced gi:8248731.
COMMENT     HTGS_phase1: Unordered fragments separated by 100 N nucleotides

```

```

contig 01 1..37893
contig 02 37994..64306
contig 03 64407..91639
contig 04 91740..94000
contig 05 94101..134454
contig 06 134555..140751
contig 07 140852..147243
Clone received from the Resource Centre of the Human Genome Project
at the Max-Planck-Institute for Molecular Genetics.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 37893: contig of 37893 bp in length
* 37894 37993: gap of 100 bp
* 37994 64306: contig of 26313 bp in length
* 64307 64406: gap of 100 bp
* 64407 91639: contig of 27233 bp in length
* 91640 91739: gap of 100 bp
* 91740 94000: contig of 2261 bp in length
* 94001 94100: gap of 100 bp
* 94101 134454: contig of 40354 bp in length
* 134455 134554: gap of 100 bp
* 134555 140751: contig of 6197 bp in length
* 140752 140851: gap of 100 bp
* 140852 147243: contig of 6392 bp in length.
FEATURES
             Location/Qualifiers
             1..147243
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="17"
             /map="17p11.2"
             /clone="PAC RPCI-1 77H15"
             /clone_lib="RPCII,3-5 Human PAC library, originating
             institute: Roswell Park Cancer Institute, creator: Pieter
             de Jong, P. Ioannou"
             /note="region between markers D17S842-D17S953"
             misc_feature
             1
             /note="T7_end:PAC RPCI-1 77H15"
             misc_feature
             147243
             /note="SP6_end:PAC RPCI-1 77H15"
BASE COUNT  39591 a 33759 c 34636 g 38657 t 600 others
ORIGIN

Query Match          3.6%; Score 56; DB 84; Length 147243;
Best Local Similarity 54.5%; Pred. No. 0.011;
Matches 134; Conservative 0; Mismatches 110; Indels 2; Gaps 1;

Qy 1149 gtggagagaattgagaacacagtggaaagtgtaaatcgtaaaagaggttgaa 1208
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10782 GAGGAGGAAGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10839

Qy 1209 gagatagaagaagagtagagaagaagtggaaggttcggtagactgaacttg 1268
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10840 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10899

Qy 1269 tccagagattggtcgaaggatttaataactgaagcaagcagagaggaattcgtgaga 1328
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10900 GAGGAGGAAGAAGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10959

Qy 1329 gagttaagtaagaagagatggagtttttgatgagatcaaaatggaaagaggtt 1388
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10960 GAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 11019

Qy 1389 gaaaaa 1394
      ||| | |
Db 11020 GAAGAA 11025

```

RESULT 11	AC009601	39987 bp	DNA	INV	20-JUN-2000	Leishmania major chromosome 35 clone L165 strain Friedlin, complete sequence.
LOCUS	AC009601.3	GI:8573033				
DEFINITION	AC009601					
ACCESSION	AC009601					
VERSION	HTG.					
KEYWORDS	Leishmania major.					
SOURCE	Leishmania major.					
ORGANISM	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.					
REFERENCE	1 (bases 1 to 39987)					
AUTHORS	Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M., Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.					
TITLE	Direct Submission					
JOURNAL	Submitted (28-AUG-1999) Seattle Biomedical Research Institution, 4 Nickerson Street, Seattle, WA 98109-1651, USA					
REFERENCE	2 (bases 1 to 39987)					
AUTHORS	Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L., McDonach,P., Ivens,A., Nguyen,D., Munden,H., Sunkin,S.M. and Stuart,K.					
TITLE	Direct Submission					
JOURNAL	Submitted (20-JUN-2000) Seattle Biomedical Research Institution, 4 Nickerson Street, Seattle, WA 98109-1651, USA					
COMMENT	On Jun 20, 2000 this sequence version replaced gi:6001967.					
FEATURES	Location/Qualifiers					
source	1..39987					
	/organism="Leishmania major"					
	/strain="Friedlin"					
	/db_xref="taxon:5664"					
	/chromosome="35"					
	/clone="L165"					
	<2..34					
	/gene="LNK1"					
gene	/note="611.6; ORFdelta; predicted using Glimmer, Testcode and CodonUsage, Glimmer predicts second ATG as start codon"					
	<2..34					
	/gene="LNK1"					
	/note="has similarity to nrk kinase, contains kinase domains"					
CDS	/codon_start=1					
	/product="Leishmania NIMA-related Kinase 1"					
	/protein_id="AAF77207.1"					
	/db_xref="GI:8573034"					
	/translation="ISKFRACGM"					
	4345..5115					
	/gene="611.8"					
gene	/note="predicted using Glimmer, Testcode and CodonUsage weak"					
	4345..5115					
	/gene="611.8"					
	/note="contains zinc finger domain, contains similarity to deoR family (bacterial regulatory proteins)"					
CDS	/codon_start=1					
	/product="611.8"					
	/protein_id="AAF77208.1"					
	/db_xref="GI:8573035"					
	/translation="MSQQQWTPDORRQOIAEMLRNKGVCIRDFLFGRCRSRPTCPYM VPATLNPVPSVCTFTQCKLRCGCSFFHGTAQLEELHASGAPVVRPQDYMIA HPAEYLNAGSIATHLISIAAPVPTPALHVVPPTYSHENAVNSFOPVLVMPQNSOAI APTFASQGRTHASPSPHAFYANAPITSONPHTSAATVLOPLHTVQPTTFQAANA LLAQQQHFAFVPLQNTLPQHQQLTQALGSRVFIHQPH"					
	11016..12407					
	/gene="611.9"					
gene	/note="predicted using Glimmer, Testcode and CodonUsage"					
	11016..12407					
	/gene="611.9"					
CDS	/note="contains ATP/GTP binding domains, has similarity to GTPases"					
	/codon_start=1					
	/product="L165.1"					
	/db_xref="GI:8573036"					
	/translation="MRATLRQCNVDFPAMRKVKRTLISAVFSGKSMETSAKKSMSREAFV DIDEKGAWYLGHWQSAANLADVKDADFILEIROARLPFTTENENIRKLTAGKPRFL IIPNKAELSNEDSNRAIOEYERNGAFALFSAARCWRDVEAVQRFTHILLPLPYK TVAHVGLVGNPNVKSTLNSLFLKLAHEYOFHREDFRSRSPETVSITPTTGKMKLV SGSEHMAECILHLRPVPISTFDCVAMICERSGTSGOTENGLDVPVRAHRRFFVHDFSDV NLGKITLDVLPRLRLRSSQSSIPLESSTGTAATDGAQAAAGDDGAYAWTHHVETSDV VERYPDHMDVLOSILKGLPHDGVPRRHQAPSVKNLSDSPDCTVISRKKGPISTRVATF DEQLKRHTRLIPGR"					
	13052..14005					
	/gene="611.10"					
gene	/note="predicted using Glimmer, Testcode and CodonUsage, Glimmer predicts second ATG as start codon"					
	13052..14005					
	/gene="611.10"					
CDS	/codon_start=1					
	/product="611.10"					
	/protein_id="AAF77210.1"					
	/db_xref="GI:8573037"					
	/translation="MANASAADELRLRTOQYLVEIDECVFVEROBELAERLSLSQL ETQKTLVNDLFVLVDVCRTYTIEDAILPALLRTLDPASSPAQQLQSRNWRDALT RLRCRGDRSDSVPLLATPRAFSRAADVPLISHDSAAAGAHVQRLLAEHEARLL DALARATARIETHAATAENGELALARMNELTRVYTSGSSMWLKAQPPPEASAAQPP EGVSSATAPFAEARALAYEKTQVALNNELALLHENNAALSRARTREVDMLKRHMAEDQ RKHEDQIAECDAVILGRMSVLEQLIQENAKLKHKLRMTAELD"					
	14801..15382					
	/gene="611.11"					
gene	/note="predicted using Glimmer, Testcode and CodonUsage"					
	14801..15382					
	/gene="611.11"					
CDS	/note="contains S-adenosylmethionine-dependent methyltransferase domain,has Blastp similarity to hypothetical protein W02B12.10 from Caenorhabditis elegans"					
	/codon_start=1					
	/product="611.11"					
	/protein_id="AAF77211.1"					
	/db_xref="GI:8573038"					
	/translation="MVNPVFVDVGCAGMLFSLAAVFPATCMGLEIRPKVVFSA QEKALALQENASLTQHYRNWPFQNMNVMKFGSCNFTKGQLSLRFLFCYPDP					

```

/protein_id="AAF77213.1"
/db_xref="GI:8573040"
/translation="MLMCSTFASLCGNPCLPDPRLWLGALYCLRLMGVACVYRDPARPS
VEVARVPRAADQOPARKACFLSCGWSARGLDSVATPLQVVSQVAAQTHRLDSDL
TWATPPPSFSLASHLSIPRYFVSSHQHRRGTHAGAVYVTTGCRVRLITFTVLVL
AALALICHRKRAYRANTVGFUHAAGAGGGERVWALDGLQHADAAAGVKQYVL
FTNEYKPADRLAESDQHLLSLVKQFISIRLLRFRFYLRPALTRWLSGDYAPRLT
LILOTFWGAALFVAVANAVPTIVETVGFVFFAYPLRLIAGCMVSYTHYPTVSS
AMQVRSEVSESHRTSPTVAMNMLRCARVYGVFSLRYRCMGFFPNVNLNNSWTO
NIVQSTFWPRACILYPPCDVAGFAAGSOPPALRNNRIVSVQCFRKNHMLQLVAFH
AAMPRLPRDAKLMVIGGARNADDRKRAQLHVRKELGIEQVELLYVNAVVAEQVEL
GKCVLGLHTRMDEHGFVILEYLAAGCIPLGRHSGVDELIDLSFDLGLAVTAEYA
AAWVEICMRLEDPRYVQFQKRGSEHKVSFSDSSFRFVELVSEYYIAC"
23148..23402
/gene="L165.2"
/notes="predicted using Glimmer, Testcode, CodonUsage"
23148..>23402
/gene="L165.2"
/notes="contains Gribkov GLNA_1 Glutamine synthetase
signature 1"
/codon_start=1
/product="L165.2"
/protein_id="AAF77214.1"
/db_xref="GI:8573041"
/translation="MSYRSSEAKKEFRKYLESTQVDALRVLVNLYEEBEKPEDPV
DIKRVIGASADYEALQOENARLRAEVELLKOLSGOQ"
24858..25517
/gene="L165.3"
/notes="predicted using Glimmer, Testcode, CodonUsage"
24858..>25517
/gene="L165.3"
/notes="contains WD40 repeats"
/codon_start=1
/product="L165.3"
/protein_id="AAF77215.1"
/db_xref="GI:8573042"
/translation="MKRVDSAKPNDFCSPVTCSDWNSDINTVACCSIDSTVTLWDV
EFGAKTKLVYHDKVDYFADIASAHTFASCAGDSVRFDFLNMDCHTLITLWQSLP
LURLAWNQDFYATFGLDSDPAVVIDMRYTPVPSOLHLPINLFWSPONQAP
NICTAGEDGLVCWEAREAKRGSILWCDEVPINNVAWRRAQNEWMIAITTSKGAQLL
PL"
26533..27297
/gene="L165.4"
/notes="predicted using Glimmer, Testcode, CodonUsage"
26533..>27297
/gene="L165.4"
/codon_start=1
/product="L165.4"
/protein_id="AAF77216.1"
/db_xref="GI:8573043"
/translation="MSLNSLTQTLVLRDMSGNVTOOKLHNRFVDFAYEAKSLVFOA
I3PAQOLVMKQYSGRIPPMHPPLGQPLMVDWSSELVELHKPDNEYQLPRRANNAYA
VNSAICCSGSPFENDHLEPDLVDFKLFKQADQADARTAFNLKHTKVPQTFILDGLM
EAPKASALYSPHNILTPAHVNNLAGTEQFLREWCRAPDGDRHRQLKLCFSLLKQOT
HLFLGTNAAPGRLELLNYAKGKNFVYAKKGMAFYVP"
27619..27753
/gene="L165.5"

```

```

Query Match 3.6%; Score 55.6; DB 4; Length 39987;
Best Local Similarity 50.0%; Pred. No. 0.011;
Matches 139; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 1113 agaacgataacacatggtctcgaacctgcgtcgtgagagaattgagaagacagt 1172
* Db 2713 AGAAGGAAACAAAAGTGCACCGCTTTCAGGCGCGGAGACGACGCGCGGATGTGG 2654
QY 1173 gaagaaggtgaagataatcgtaaaagaggttgaagagataagaagaagatagagaag 1232
+Db 2653 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2594
QY 1233 gaagtggagaagtcggtgagctgagatgaactgttccagagattggtcgaagatt 1292
Db 2593 GAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2534

```

```

QY 1293 aatgaactgaagcaagcagcagagaatttcgtgagagagtttaagtaagaagagatggag 1352
Db 2533 GAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2474
QY 1353 ttttggatgagatcaaaatggaagcaagtggatgtga 1390
Db 2473 GAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2436

RESULT 12
AC005804 42160 bp DNA INV 31-MAY-2000
LOCUS Leishmania major chromosome 35 clone 6.1.1.1 strain Friedlin,
DEFINITION complete sequence.
ACCESSION AC005804
VERSION AC005804.2 GI:6006841
KEYWORDS Htg.
SOURCE Leishmania major.
ORGANISM Leishmania major.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 42160)
AUTHORS Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrook,M.,
Cawthra,J., Marsolini,F., McDonagh,P., Stuart,K.D., Cunningham,M.
and Beverley,S.
TITLE Direct Submission
JOURNAL Submitted (09-OCT-1998) Seattle Biomedical Research Institution, 4
REFERENCE 2 (bases 1 to 42160)
AUTHORS Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrook,M.,
Cawthra,J., Marsolini,F., McDonagh,P., Stuart,K.D., Cunningham,M.
and Beverley,S.
TITLE Direct Submission
JOURNAL Submitted (02-OCT-1999) Seattle Biomedical Research Institution, 4
REFERENCE 3 (bases 1 to 42160)
AUTHORS Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrook,M.,
Cawthra,J., Marsolini,F., McDonagh,P., Stuart,K.D., Cunningham,M.
and Beverley,S.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2000) Seattle Biomedical Research Institution, 4
COMMENT Nickerson Street, Seattle, WA 98109-1651, USA
FEATURES On Oct 2, 1999 this sequence version replaced gi:3723957.
source
1. 42160
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/chromosome="35"
/clone="6.1.1"
/notes="Also known as 611"
535..>2694
/gene="511.9"
535..>2694
/gene="511.9"
/notes="contains a Proline-rich region (PRO_RICH),
contains recA family and SURF1 domains (Pfam)"
/codon_start=1
/product="511.9"
/protein_id="AAF73077.1"
/db_xref="GI:8122319"
/translation="MWAVAQLEQLNDLALNRASAKATDSFADTSATSAIRANEIIR
DVKPDVAQENITLTIEAKAARCYHPPPIPAVLTKGSSWEALARCMDLYVADD
YLASHPPNHYGTIEARTEMQLQOIVRISDFVKSFIINAVORSKYSPSGADGTG
OAGSLSVAPMDSALTAPTANRLIRNEAALQGVQIVORLGENFRTDILRKDVRL
LEKLVAVQAQDFGARDEETGRPLTSRSSLLPVLYKHQGHHPHLLRVSKSARELY
TDACALQRYVLTFFLEDDYAVADIPSELATVVELYVRAIKVIQDLTSTFFADPKLF
VDSRGQGLFSTPYKPRFYDFITGLDFMEEFKWKMLSKSIPGENSEDFIDHVDSDVN
FYIRLELDLQYNAKALEKESKDFEYMRRTWFPSPVDCVTGHTSTNTVLYFHKTL
FTYFSGSLRVLYGSVIGANADYEQVEDYITRGVAGVDDQLQVLAAABALQOEA
LAKRNHTKSNLNLTSVRLSDVDFMINLNCFLDNTYRREACTFRLAAATPPPKS
SCEGASARKATAAPPPDPPLSLQFDMNLNERYVEAFGASWHKCFPSIKGQSDLLS

```



```

REFERENCE 1 (bases 1 to 2813)
AUTHORS Nimzyk,R., Schondorf,T. and Hachtel,W.
TITLE In-frame length mutations associated with short tandem repeats are
located in unassigned open reading frames of Oenothera chloroplast
DNA
JOURNAL Curr. Genet. 23 (3), 265-270 (1993)
MEDLINE 93169690
REFERENCE 2 (bases 1 to 2813)
AUTHORS Hachtel,W.
TITLE Direct Submission
JOURNAL Submitted (18-FEB-1992) W. Hachtel, Botanisches Institut der Uni
Bonn, Kirschallee 1, 5300 Bonn 1, FRG
FEATURES
source 1..2813
/organism="Oenothera odorata"
/organella="plastid"
/db_xref="taxon:3951"
<1..2166
/note="ORF2280"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAA45898.1"
/db_xref="GI:14335"
/db_xref="SWISS-PROT:P31568"
/translation="EFESTILGPSVPDLVALTNEALSISITQKKSIIIDTTIRYALHR
KTDLEADRNLSPAKHGTLFYQVGRFAHTVLLRNCISIDPSIYIKKNCLEAGDSSL
YKMYFELGTSMKKLTILLYLLTCSAGSIAQDLSPGPDEONLITSYGLVENDSDLVH
GLSDIYHGLLEGLGSSPTTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTE
EEVGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTE
EEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTE
EEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTE
DEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTE
DEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTEVEGTE
DDEVDVFAKMLEDLSELVWSPRIWHPWDFLLDCEAEIPAEIPEDEDELPEADALE
TEVAVNGVEEGEADDEEDVLEAOQEDLEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
LHEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE
EISKELLRLNPKTRDKPRQWTKKKQDKHYELLDRQRLITRSLSKSNGF
FRSNTPSESYQYLSNLFNSRRLDQMTKTFRRKMAFFG"
2279..2422
/number=1
join(2279..2422,2705..2803)
/note="ORF80"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAA45899.2"
/db_xref="GI:13513074"
/db_xref="SWISS-PROT:P31566"
/translation="METPVSSIFWTLAPRNNMLLKDGRIELDQNTMYGWYELPKQE
FLNSVRTQIFTKKCYIPFRIGPERRKAGMPTGV"
2423..2704
/number=1
2705..2803
/number=2
BASE COUNT 989 a 444 c 798 g 582 t
ORIGIN

Query Match 3.5%; Score 54.8; DB 14; Length 2813;
Best Local Similarity 50.0%; Pred. No. 0.012;
Matches 164; Conservative 0; Mismatches 162; Indels 2; Gaps 1;

Qy 1080 aaagcataggagagacttcagacattcattagacgataacacatgtgctcgtgaa 1139
Db 894 AGAAGAGGAAGTAGAGGGACAGAGAGGAGTAGAGGGACAGAGAGATGAGG--AAGTA 951

Qy 1140 cctgcgctcgtggagagaatttgagaagacagtggaagaaggtagaataatcgtaaaa 1199
Db 952 GAAGGGACAGAGAGGAAGTAGAGGGACAGAGAGGAGGAGTAGAGGGACAGAGATGAG 1011

Qy 1200 gaggttgagagatagaagaaggttagagaagaagtagagaaggtcggttaggactgag 1259
Db 1012 GAAAGTAGAGGGACAGAGAGGAAGTAGAGGGACAGAGAGGAGGAGAGTAGAGGGACAGAA 1071

```

```

Qy 1260 atgacctgttccagagatiggtcgaagatttaataatgaactgaagcaagacgagagaaat 1319
Db 1072 GAGGAAGTAGAAGGGACAGAGAGGAAAGTAGAAGGGACAGAGAGGAGTAGAAGGGACA 1131
Qy 1320 ttctgagagagagtttaagttaaaagaagagatggagtttttttgatgagatcaaaatggaaagca 1379
Db 1132 GAAGAGGAAGTAGAAGGGACAGAGAGGAAAGTAGAAGGGACAGAGAGGAGGAGTAGAAGGG 1191
Qy 1380 agtgaggttgaaaaaattgtttgggaaag 1407
Db 1192 ACAAGATGAGGAAGTAGAAGGGACAG 1219

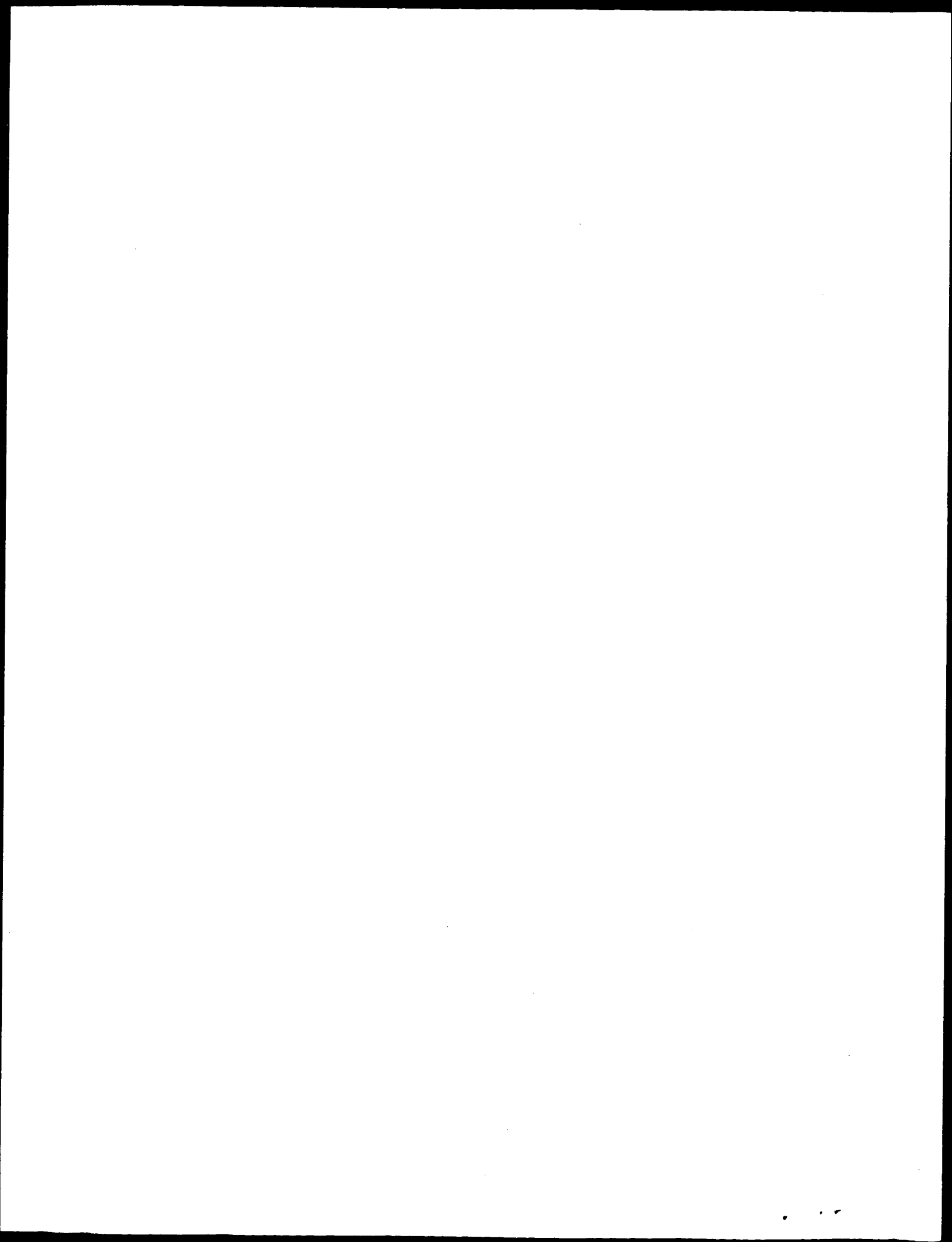
```

Search completed: November 6, 2001, 03:10:44
Job time: 12223 sec

Wed Nov 7 09:24:02 2001

us-09-075-375a-5.rge

Page 19



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2001, 02:55:53 ; Search time 313.48 Seconds
(without alignments)
3114.671 Million cell updates/sec

Title: US-09-075-375A-5

Perfect score: 1555

Sequence: 1 ccacgcgtccgcttggtgt.....ggataaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_0601.*
1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1553.4	99.9	1555	18	Arabidopsis violax
2	624.2	40.1	1589	18	Tobacco violaxanth
3	608.6	39.1	1981	18	Romaine lettuce v1
4	104	6.7	936	22	Oligonucleotide D1
5	104	6.7	936	22	Oligonucleotide D1
6	104	6.7	936	22	Oligonucleotide D1
7	104	6.7	936	22	Oligonucleotide D2
8	104	6.7	936	22	Oligonucleotide D2
9	104	6.7	936	22	Oligonucleotide D1
10	102.8	6.6	936	22	Oligonucleotide D1
11	102.8	6.6	936	22	Oligonucleotide D1

c 12	102.8	6.6	936	22	AAF58257	Oligonucleotide D1
c 13	102.8	6.6	936	22	AAF58259	Oligonucleotide D2
c 14	102.8	6.6	936	22	AAF58262	Oligonucleotide D1
c 15	102.8	6.6	938	22	AAF58255	Murine LOBO homolo
c 16	50.2	3.2	49999	20	AAZ23891	Murine LOBO homolo
c 17	50.2	3.2	49999	20	AAZ23896	Oligonucleotide D1
c 18	50	3.2	244	22	AAF58238	Oligonucleotide D1
c 19	46.6	3.0	2169	21	AAZ50687	Arabidopsis thalia
c 20	46.6	3.0	2638	21	AAZ47245	Arabidopsis thalia
c 21	46.4	3.0	244	22	AAF58238	Oligonucleotide D1
c 22	46.4	3.0	3211	18	AAZ89346	Human p160 cDNA 16
c 23	46.4	3.0	3901	18	AAZ89345	Human p160 cDNA 16
c 24	45.6	2.9	1067	21	AAZ44224	Arabidopsis thalia
c 25	45.4	2.9	29392	19	AAZ15422	Mouse poly Ig rece
c 26	44.4	2.9	7215	20	AAZ13039	Enterococcus faeca
c 27	44.2	2.8	2277	19	AAZ05370	Human telomerase p
c 28	44	2.8	1236	21	AAZ02163	Human colon cancer
c 29	44	2.8	1664976	19	AAZ21209	Methanococcus jann
c 30	43.6	2.8	2277	19	AAZ13834	Homo sapiens ambig
c 31	43.2	2.8	3489	21	AAZ30290	Kaposi's sarcoma-a
c 32	43.2	2.8	32207	20	AAZ73805	KSHV LUR DNA (nucl
c 33	43.2	2.8	137507	19	AAZ19941	KSHV long unique c
c 34	42.6	2.7	1798	18	AAZ51756	DAX-1 gene promote
c 35	42.4	2.7	5935	21	AAZ28818	Murine T cell indu
c 36	42.4	2.7	6888	21	AAZ59599	DNA encoding a cyt
c 37	42.2	2.7	2643	14	AAZ39212	CENP-B cDNA. HOMO
c 38	42.2	2.7	3717	21	AAZ64660	DNA encoding centr
c 39	42	2.7	1607	17	AAZ35277	Chemokine receptor
c 40	42	2.7	1677	21	AAZ58872	DNA encoding a hum
c 41	42	2.7	2324	21	AAZ36968	Arabidopsis thalia
c 42	41.6	2.7	1313	21	AAZ33413	Arabidopsis thalia
c 43	41.4	2.7	1835	20	AAZ91242	T. gondii immunoge
c 44	41.2	2.6	1559	18	AAZ91855	DUB-1 enhancer/pro
c 45	41	2.6	4677	21	AAZ70259	Plasmodium falcipa

ALIGNMENTS

RESULT 1

AAZ66243
ID AAT66243 standard; cDNA; 1555 BP.
XX
XX AAT66243;
XX
XX 28-JUL-1997 (first entry)
XX
XX Arabidopsis violaxanthin de-epoxidase cDNA.
DE
DE Violaxanthin de-epoxidase; VDE; light; photosensitivity;
KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
KW xanthophyll; lettuce; ss.
XX
XX Arabidopsis thaliana cv. columbia.
XX
XX Key Location/Qualifiers
FH CDS 45..1433
FT /*tag= a
FT transit_peptide 45..383
FT /*tag= b
FT mat_peptide 384..1430
FT /*tag= c
XX
XX WO9717447-A2.
PN
XX
XX 15-MAY-1997.
PD
XX
XX 07-NOV-1996; 96WO-US18291.
PF
XX
XX 06-AUG-1996; 96US-0023502.
PR
XX
XX 07-NOV-1995; 95US-0006315.
XX
XX (CALJ) CALGENE INC.

XX

P1 Bugos RC, Rockholm DC, Yamamoto HY;

XX
DR WPI: 1997-281036/25.
DR P-PSDB; AAW09876.XX
PT DNA encoding plant violaxanthin de-epoxidase - used to modify the
PT sensitivity of a plant to lightXX
PS Disclosure; Fig 3; 41pp; English.XX
CC A cDNA clone (AA066243) codes for Arabidopsis violaxanthin
CC de-epoxidase (VDE) (AAW09876), an enzyme that catalyses the
CC de-epoxidation of violaxanthin to zeaxanthin and antheraxanthin.
CC VDE nucleic acids (see also AA066241-42), in sense or antisense
CC orientation, can be used in genetic constructs, pref. also contg. a
CC plastid translocation sequence, to modify VDE levels in plants.
CC Increased levels result in the plant being tolerant of increased
CC light and therefore more productive and/or more resistant to
CC disease. Underexpression of VDE increases photosynthetic
CC efficiency under low light. The photosensitivity of a range of
CC crops, trees and ornamentals can be modified.XX
SQ Sequence 1555 BP; 485 A; 286 C; 375 G; 409 T; 0 other;Query Match 99.9%; Score 1553.4; DB 18; Length 1555;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1554; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ccacgcgtcggcttggtggtgggaagattagatagtggaagaatggcagtagctaac 60
 Db 1 ccacgcgtcggcttggtggtgggaagattagatagtggaagaatggcagtagctaac 60

QY 61 attgtttcaactcacttgcacgtgattcgatttctcgaatttctcgaatgatgtattg 120
 Db 61 attgtttcaactcacttgcacgtgattcgatttctcgaatgatgtattg 120

QY 121 gtaggcttgccattacaagaagaggatcaatggcactttctgtcgaattttacctc 180
 Db 121 gtaggcttgccattacaagaagaggatcaatggcactttctgtcgaattttacctc 180

QY 181 caatccaaagtctgattccaaacactggtggagatccacgtcttattctgcat 240
 Db 181 caatccaaagtctgattccaaacactggtggagatccacgtcttattctgcat 240

QY 241 tcaggctcaggattctctaaggggatattgacattgtgccattaccatcaagaatgagc 300
 Db 241 tcaggctcaggattctctaaggggatattgacattgtgccattaccatcaagaatgagc 300

QY 301 tgaagagctgacccgtccgcgtgtgtcctaaactcgtgggtgttttagctgcgcttcc 360
 Db 301 tgaagagctgacccgtccgcgtgtgtcctaaactcgtgggtgttttagctgcgcttcc 360

QY 361 ttattgttccattcgcagatgcagttgatgcacttgaacttgcactgtcattgaagg 420
 Db 361 ttattgttccattcgcagatgcagttgatgcacttgaacttgcactgtcattgaagg 420

QY 421 gatcgaggtatagaactcgcgaagtgcattgccaaacctgcctgtgcgccaatgtcgcgt 480
 Db 421 gatcgaggtatagaactcgcgaagtgcattgccaaacctgcctgtgcgccaatgtcgcgt 480

QY 481 gccctcagacctgcaataaccctccagatgaaccagtgccagattaaatgtgggagtc 540
 Db 481 gccctcagacctgcaataaccctccagatgaaccagtgccagattaaatgtgggagtc 540

QY 541 tgtttgagaacagtggtgttgatagttcaacgagtgctgtgtcgcgagaaaaagtgtg 600
 Db 541 tgtttgagaacagtggtgttgatagttcaacgagtgctgtgtcgcgagaaaaagtgtg 600

QY 601 ttccctagaaaactgatctcgagaatttctgcgccagaccctctctgtttgttacaga 660
 Db 601 ttccctagaaaactgatctcgagaatttctgcgccagaccctctctgtttgttacaga 660

QY 661 acttcaacatctcggacttttaacggaagtgttacattacaagtggcttgaatccaacct 720
 Db 661 acttcaacatctcggacttttaacggaagtgttacattacaagtggcttgaatccaacct 720

QY 721 ttgatgcttcgactgcagctgcagtgatgatttccacagaagtgacacaagtgttg 780
 Db 721 ttgatgcttcgactgcagctgcagtgatgatttccacagaagtgacacaagtgttg 780

QY 781 gaaacatctcttgagaataaagaccctacagatgacatgattcttactagtcagcgtac 840
 Db 781 gaaacatctcttgagaataaagaccctacagatgacatgattcttactagtcagcgtac 840

QY 841 aaaaattcgtcaagatcttaacccaacctggttctctcaaatcatgacaacgagttacc 900
 Db 841 aaaaattcgtcaagatcttaacccaacctggttctctcaaatcatgacaacgagttacc 900

QY 901 ttcaactcaagatgactgtgtatctctgtctcaaatgagagatataaacctgaagact 960
 Db 901 ttcaactcaagatgactgtgtatctctgtctcaaatgagagatataaacctgaagact 960

QY 961 atatttctatactacgtggcgaaacgatccttgggatgatatgtgtgcaagtgtg 1020
 Db 961 atatttctatactacgtggcgaaacgatccttgggatgatatgtgtgcaagtgtg 1020

QY 1021 tatacagagaagtctgtattaccccaatgacattaccagaactcgaaaaagcagcaa 1080
 Db 1021 tatacagagaagtctgtattaccccaatgacattaccagaactcgaaaaagcagcaa 1080

QY 1081 aaagcataggcagagacttcagacatttcattagaacgagatacacatgtgctcctgaac 1140
 Db 1081 aaagcataggcagagacttcagacatttcattagaacgagatacacatgtgctcctgaac 1140

QY 1141 ctgcgtcgtggagagaatttgaaagacagtggaaagagtgaagataatcgtataaag 1200
 Db 1141 ctgcgtcgtggagagaatttgaaagacagtggaaagagtgaagataatcgtataaag 1200

QY 1201 aggttgaagagatagaagaagaggtagagaagaagtggagaaggtcggttaggactgaga 1260
 Db 1201 aggttgaagagatagaagaagaggtagagaagaagtggagaaggtcggttaggactgaga 1260

QY 1261 tgacctgttccagagattgctgaaagatttaatagaactgaagcagcagagagaatt 1320
 Db 1261 tgacctgttccagagattgctgaaagatttaatagaactgaagcagcagagagaatt 1320

QY 1321 tcgtgagagagtttaagtaaaagaagagatggagtttttggatgagatcaaaatggaagcaa 1380
 Db 1321 tcgtgagagagtttaagtaaaagaagagatggagtttttggatgagatcaaaatggaagcaa 1380

QY 1381 gtgaggttgaaaaattgtttgggaagccttgcgaatcagaagaaggtcaggtagaacaag 1440
 Db 1381 gtgaggttgaaaaattgtttgggaagccttgcgaatcagaagaaggtcaggtagaacaag 1440

QY 1441 aaccacattgttgtcaaaactattatacactgtgttcgtgttcattataaagtaata 1500
 Db 1441 aaccacattgttgtcaaaactattatacactgtgttcgtgttcattataaagtaata 1500

QY 1501 tttttgtacagtcacatcatcttcatacaaatgttgataaaaaaataaaaaa 1555
 Db 1501 tttttgtacagtcacatcatcttcatacaaatgttgataaaaaaataaaaaa 1555

RESULT 2

AAT66242

ID AAT66242 standard; cDNA; 1589 BP.

XX AC AAT66242;

XX DT 28-JUL-1997 (first entry)

XX DE Tobacco violaxanthin de-epoxidase cDNA.

XX VDE; light; photosensitivity;

XX VDE; light; photosensitivity;

KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 XX xanthophyll; lettuce; ss.

OS Nicotiana tabacum L. cv. xanthi.

XX Key Location/Qualifiers
 FH 42..1478
 FT /*tag= a
 FT transit_peptide 42..443
 FT /*tag= b
 FT mat_peptide 444..1475
 FT /*tag= c

WO9717447-A2.

15-MAY-1997.

07-NOV-1996; 96WO-US18291.

06-AUG-1996; 96US-0023502.

07-NOV-1995; 95US-0006315.

(CALJ) CALGENE INC.

Bugos RC, Rockholm DC, Yamamoto HY;

WPI; 1997-281036/25.

P-PSDB; AAW09875.

DNA encoding plant violaxanthin de-epoxidase - used to modify the
 sensitivity of a plant to light

Disclosure; Fig 2; 4lpp; English.

CC A cDNA clone (AAT66242) codes for tobacco violaxanthin de-epoxidase
 (VDE) (AAW09875), an enzyme that catalyses the de-epoxidation of
 violaxanthin to zeaxanthin and antheraxanthin. VDE nucleic acids
 (see also AAT66241, AAT66243), in sense or antisense orientation, can
 be used in genetic constructs, pref. also contg. a plastid
 translocation sequence, to modify VDE levels in plants. Increased
 levels result in the plant being tolerant of increased light and
 therefore more productive and/or more resistant to disease.
 CC Underexpression of VDE increases photosynthetic efficiency under
 CC low light. The photosensitivity of a range of crops, trees and
 CC ornamentals can be modified.

XX Sequence 1589 BP; 500 A; 261 C; 367 G; 461 T; 0 other;

Query Match 40.1%; Score 624.2; DB 18; Length 1589;

Best Local Similarity 74.3%; Pred. No. 1.7e-162;

Matches 820; Conservative 0; Mismatches 268; Indels 15; Gaps 2;

QY 364 ttgttccatctgcagatgcagttgatgcacttaaaattgtgcatgtcttttaagggat 423

DB 424 ttgttccaaagctgatgcggttgatgctcgaagctgtactgtcttactgaagat 483

QY 424 gcagatagaactgcgaagtgcaattgccaccctgctgtgcagccaattgcgctgcc 483

DB 484 gcaggttagactgcgaagtgcaattgccaccctgcatgtgcagctaatgttgcctgtc 543

QY 484 ttccagactgcaataaccctgcagatgaaccgagtgccagattaaattgtgggactgt 543

DB 544 tccagactgcaataaccgagtgcaacggaatgtcagataaaattgtgtattgt 603

QY 544 ttgagaacagtgtgttgatgagttcgaacgagtgctgtgtcgaagaaaaagtgtttc 603

DB 604 ttgaaaaacagtgtcgtagacgagttcaatgagtgctgcagctcccgaaagaaattgtac 663

QY 604 ctgaaaaatctgactcggagaatttctgccacagaccttctgtctttctacagact 663

DB 664 ctctgaaatctgactgtgtgacttctgtactgtacctgtcccgactgtctgtccagagt 723

QY 664 tcaacatctcggacttttaacgggaagtgtgtacattacaagtgtgtgaatcaaacctttg 723
 DB 724 ttgacatgaagatatttagcgggaattgttactactcgggtttgaatccactttt 783
 QY 724 atgccttcagctccagctgcagtgattccacacagaagggtgacaacaagctctgttgaa 783
 DB 784 atgcttttgatgcgaattgcagtgatttccatcacagaa---gaaacaaacttgcggga 840
 QY 784 acatctcttggagaataaagaccctagacagtggtattcttactaggtcagcgtacaaa 843
 DB 841 atttatcttgagaatagctacacctgattgaggtatttttactcgatcagcgggtgcaa 900
 QY 844 aattcgtcgaagtactcaaacacctgtgttctctcaaatcatgacaacagtagaccttc 903
 DB 901 aattcgtcgaagtactcaaacacctgtgttctctcaaatcatgacaacagtagaccttc 960
 QY 904 actatcaagatgactggtatatctctcatcaaaagatagagaataaacctgaagactata 963
 DB 961 tctaccaagatgactggtatatctctcatcaaaagatagagaataaacctgaagattaca 1020
 QY 964 tattgtatactaccgtggcgcaaacgagtgttggtgagtgatgtgtgtcagttgtat 1023
 DB 1021 tattgtgtactataaaggcgcaaaatgatgcagtgatgtgtgtgtgtgtgtgtt 1080
 QY 1024 acacgagaagtctgtattaccatagcattatcacccgaactcgaagaaagcagcaaaaa 1083
 DB 1081 acacaagaagtgcagtttgcctgaaagcattatcacccgaggtgcgaacccgagctcaa 1140
 QY 1084 gataggcagagacttcagcacattcattagaaacgagataaacacatgtgtcctgaacctg 1143
 DB 1141 aagttggcggtgatttcaacacattcaataaaacagacatacatatgtggccctgaacctc 1200
 QY 1144 cgtcgtgcagagaattgagaagacagtggaagaggtgaaaggtataatcgttaaaagag 1203
 DB 1201 ccttctgtgaggggttggaagaaagtggaagaaagtggaagaaagcagatcataaaagaag 1260
 QY 1204 ttgaaagatagaagaagaggttagagaagagtggaagaggtcggtaggactgagatga 1263
 DB 1261 ttgaggagatagaagaagagtagagaaggtgagagaaggtgagataaagaag-----tca 1308
 QY 1264 cctgttcagagattgctgaaggtatttaagaaactgaactgaacgaagcagaggaatttcg 1323
 DB 1309 ccttattcagtaaaactgtttgaaggttttaagagctccacagagatgaaggaactctct 1368
 QY 1324 tgagagagttaagttaaagaagagatgagtttttttgatgagatcaaaatgggaagcaagt 1383
 DB 1369 taagagagctgagcaagaagaagaatgagttgttgatgacttaaaatgggaagcaactg 1428
 QY 1384 aggttgaataattgtttgggaagcctttgccaatcagggaaggtcaggtgagaacagaagac 1443
 DB 1429 aggtagaataaactttttggcggtgttaccataaaggaaatttaaggtaatttttaa 1488
 QY 1444 caccattgtgtacaactatat 1466
 DB 1489 aactatacaatatactacat 1511

RESULT 3

AAT66241

ID AAT66241 standard; cDNA; 1981 BP.

XX

AC AAT66241;

XX

DT 28-JUL-1997 (first entry)

XX Romaine lettuce violaxanthin de-epoxidase cDNA.

DE Violaxanthin de-epoxidase; VDE; light; photosensitivity;

XX photoprotection; transgenic plant; zeaxanthin; antheraxanthin;

KW xanthophyll; lettuce; ss.

XX Lactuca sativa L. cv. romaine.

XX


```
RESULT 4
AAF58252
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 6.7%; Score 104; DB 22; Length 936;
Best Local Similarity 0.6%; Pred. No. 6.8e-19;
Matches 5; Conservative 467; Mismatches 302; Indels 0; Gaps 0;

Qy 782 aaacatctcttgagataaagaccctagacagtggttcttactagtcagccgtaca 841
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 60

Qy 842 aaattcgtcaagatactcaaccacctggtgttctctacaatcagacacagtagacct 901
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 120

Qy 902 tcaatcaagatgactggtatctctgtcatcaagaatagagataaaccctgaagacta 961
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
121 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 180

Qy 962 tatattgtatactacogtggcgaaacagatcgttggtggtatggtggtgagttgt 1021
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
181 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 240

Qy 1022 atacacgagaagtctgtattaccaatagcattataccagaactcgaagacagcaaa 1081
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
241 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 300

Qy 1082 aagcataggcagagacttcagacacatttattagacggtataacacatgtgtcctgaacc 1141
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
301 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 360

Qy 1142 tgcgtcgtgagagaattgagaagacagtggaagaaggtgaaagataatcgtaaaaga 1201
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
361 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 420

Qy 1202 ggttgaagagatagaagaagaggtagagaagaagtgagaggtcgtgtaggactgagat 1261
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
421 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 480

Qy 1262 gacctgttccagagatggtgctgaaggatttaatagaactgaagaagacgaggaattt 1321
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
481 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 540

Qy 1322 cgtgagagagtttaagtaaaagagagatgggtttttggtgagatcaaaatggaagcaag 1381
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
541 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 600

Qy 1382 tgaggtgaaaaattgtttgggaaagctttccaatcaggaaggtcaggtagaacaaga 1441
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
601 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 660

Qy 1442 accacattgtgtacaaactatattacatactgtgttcgttcataataagtaatat 1501
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
661 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 720

Qy 1502 tttgtacacagtcacatccattccatacaaatggataaaaaaa 1555
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
721 www.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww.ww 774

RESULT 5
AAF58254
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
```



```

XX 24-APR-2001 (first entry)
XX Oligonucleotide D1875.
XX Electron-transfer group; ETM; mismatch; genotyping;
XX gene expression; ss.
XX Synthetic.
XX WO200107665-A2.
XX 01-FEB-2001.
XX 26-JUL-2000; 2000WO-US20476.
XX 26-JUL-1999; 99US-0145695.
XX 17-MAR-2000; 2000US-0190259.
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX Umek RM;
XX WPI; 2001-159728/16.
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface
XX Example 6; Page 127; 159pp; English.
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 6.6%; Score 102.8; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 1.5e-18;
Matches 8; Conservative 472; Mismatches 314; Indels 0; Gaps 0;

Qy 762 ggtgacacagctgttggaacatctcttgagataaagacccttagacagtgattc 821
Db ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
794 GGGGGCGCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 822 ttactaggtcagcgtacaaaattcgtgcaagatccttaaccaacctggtctctac 881
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
734 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 882 aatcatgacacagtagtaccttcaactcaagatgactggtatctctgtcatcaagata 941
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
674 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 942 gagaataaacctgaagactatattgttatactaccgtggcgaaacagatgcttggat 1001
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
614 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1002 ggatagtggtgctgattgtatcacagagaagttctgtattaccatagcattaccca 1061
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
554 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1062 gaactcgaataaacagcaaaaagcataggcagagacttcagcacattcattagaacgat 1121
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
494 GWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1122 aacacatggtgctgaacctgcgctcgtggagagaattgagaagacagtggaagaagt 1181
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
434 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 1182 gaagdataatogtaaaagaggttgaaagagatagaagaagaggttagaag 1241
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 374 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1242 aaggtcgttagactgagatgacctgttccagagatgctggaaggatttaactgaactg 1301
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1302 aagcaagacgagagaatttcgtgagagagtttaagtaagaagagatggagttttggat 1361
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1362 gagatcaaatggaagcaagtgtggttgaaaattgttgggaagcttgcacatcagg 1421
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1422 aaggtcaggtagaacaagaacacacacattgtgtacaaactatatatacatcactgtgtt 1481
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 134 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1482 cgttcataataagtaataattttgtacacagtcacatccatccataacaattggataa 1541
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1542 aaaaaaaaaaaa 1555
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 14 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 2001, 02:51:40 ; Search time 166.68 Seconds
(without alignments)
2112.871 Million cell updates/sec

Title: US-09-075-375A-5
Perfect score: 1555
Sequence: 1 ccacgcgtccggttggtgt.....ggataaaaaaaaaaaaaa 1555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listed first 45 summaries

Database : Issued Patents.NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	1555	100.0	1555	3	US-08-747-574-3
2	624.2	40.1	1589	3	US-08-747-574-2
3	611.8	39.3	1981	3	US-08-747-574-1
4	92.4	5.9	7218	1	US-08-232-463-14
5	46.4	3.0	3211	2	US-08-574-959A-8
6	46.4	3.0	3211	4	US-09-357-014-8
7	46.4	3.0	3901	2	US-08-574-959A-6
8	46.4	3.0	3901	4	US-09-357-014-6
9	43.8	2.8	289	4	US-09-007-005-17
10	43.8	2.8	289	4	US-09-244-796-17
11	43.6	2.8	2277	1	US-08-676-967-2
12	43.6	2.8	2277	1	US-08-676-974-2
13	43.6	2.8	2277	2	US-09-098-487-2
14	43.2	2.8	3489	2	US-08-728-323A-1
15	43.2	2.8	32207	2	US-08-770-379-20
16	43.2	2.8	32207	4	US-08-757-669A-20
17	42.4	2.7	5935	4	US-09-178-973B-17
18	42	2.7	1607	3	US-08-875-573-19
19	42	2.7	1695	4	US-09-232-878-1
20	41.2	2.6	1559	4	US-09-019-095A-7
21	40.2	2.6	9636	1	US-08-323-170B-1
22	39.6	2.5	53526	3	US-08-658-136-2
23	39.6	2.5	53577	3	US-08-658-136-1
24	39	2.5	2188	1	US-07-865-662F-10
25	39	2.5	2188	4	US-08-374-219B-10
26	39	2.5	5361	4	US-08-973-462-2
27	39	2.5	6152	4	US-08-973-462-1

Sequence 22, Appli
Sequence 1, Appli
Sequence 101, App
Sequence 209, App
Sequence 3, Appli
Sequence 3, Appli
Sequence 6, Appli
Sequence 16, Appli
Sequence 184, App
Sequence 8, Appli
Sequence 5, Appli
Sequence 101, App
Sequence 8, Appli
Sequence 3, Appli
Sequence 6, Appli
Sequence 1, Appli
Sequence 1, Appli

28 38.6 2.5 494 2 US-08-332-766A-22
29 38.2 2.5 1505 1 US-07-915-246-1
30 38.2 2.5 43795 3 US-08-742-185-101
31 37.6 2.4 51259 3 US-08-781-891-209
32 37.4 2.4 277 4 US-09-007-005-3
33 37.4 2.4 277 4 US-09-244-796-3
34 37 2.4 18443 4 US-09-078-294-6
35 36.8 2.4 634 3 US-08-981-094-16
36 36.6 2.4 1669 4 US-09-461-697-184
37 36.4 2.3 295 4 US-09-018-584A-8
38 36.4 2.3 3238 4 US-08-123-934A-5
39 36.4 2.3 3238 5 PCT-US94-10080-5
40 36.4 2.3 43795 3 US-08-742-185-101
41 36.2 2.3 19056 4 US-09-372-032-8
42 35.8 2.3 9009 1 US-07-864-004B-3
43 35.8 2.3 9009 1 US-08-251-937A-3
44 35.8 2.3 9009 1 US-08-212-133A-1
45 35.8 2.3 9009 1 US-08-474-503-1

ALIGNMENTS

RESULT 1
US-08-747-574-3
; Sequence 3, Application US/08747574
; Patent No. 6015939
; GENERAL INFORMATION:
; APPLICANT: CALGENE, INC.
; TITLE OF INVENTION: PLANT VDE GENES AND
; METHODS RELATED THEREO
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08747,574
; FILING DATE: No. 6015939ember 7, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/023,502
; FILING DATE: August 6, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/006,315
; FILING DATE: No. 6015939ember 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: 119-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1555
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
US-08-747-574-3

Query Match	100.0%;	Score 1555;	DB 3;	Length 1555;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1555;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	ccacgcgtccgcgtctgtggtgggaagattagatagatgtgaagaatggcagtagctacac	60	
Db	1	CCACGCGTCCGGCTTGGTGTGGGAGAGATTAGATAGTGTGAAGATGGCAGTAGCTACAC	60	
Qy	61	attgtttcaacttcaaccttgcatagcacgtattcattcgatttttctcaagtgaatgattg	120	
Db	61	ATTGTTTCACTTCACTTGTGTCATGACCGGTATTTCGATTTTCTCAAGTGATGATGATTG	120	
Qy	121	gtagccttgccattacaagaagaagagatcaatggcaactttctgtctcaagaattttacctc	180	
Db	121	GTAGCCTTGGCATTTACAAAGAAAGAGATCAATGGCACTTTCTTGTGTCGAAGATTTTACCTC	180	
Qy	181	caatccaaagtctgatactcagaaacactgggtgggagatcctcagctcctttatctgcat	240	
Db	181	CAATCCAAAGTGTGATCTCAGAACAACTGGTGGGAGATCCTCAGCTCTTTATCTGGCAT	240	
Qy	241	tcaagtcaggattctctaaaggggatatttgacaattfgcatttaccatacaagaatgagc	300	
Db	241	TCAGGTCAAGATTTCTTAAAGGGGATATTTGACATGTGCGCATTACCATCAAGAATGAGC	300	
Qy	301	tgaagagctgacgcgtccgcgtgttgcataaactcgtgggtgttttagcttgcgcgttcc	360	
Db	301	TGAAAGAGCTGACCGTCCGCTGTTGCTTAAACTCGTGGGTGTTTAGTTCGCGGTTC	360	
Qy	361	ttatgttccatctgcagatgcagttgtatgcacttaaaacttgtcatgttattgaagg	420	
Db	361	TTATGTGTTCATCTGCAGATGCAAGTTGATGCACTTAAAACTTTGTGCATGCTTATTGAAGG	420	
Qy	421	gatcagagatagaactcgaactcgaagtcatctgcaaacctgcctgtcacgcaaatgtccgt	480	
Db	421	GATGCAGGATAGAACTCGCAAACTGCATTTGCCAACCTGCTGTGCACCAATGTGCGGT	480	
Qy	481	gcctcagacactgcaataaocgtccagatgaaacocagtgccagattaaatgtggggatc	540	
Db	481	GCCTTCAGACCTGCAATTAACCGTCCAGATGAACCGAGTGCCAGATTAAATGTGGGGATC	540	
Qy	541	tgttgagaacagtgtgtgtgatgagttcaacgagtgctgtgtcgaagaaaaagtgtg	600	
Db	541	TGTTTGAGAACAGTGTGTTGATGAGTTCAACGAGTGTCTGTGTGCGAAGAAAAGTGTG	600	
Qy	601	ttcctagaaaactgatctcgcggagaatttccctgccccagacccctctgtctctgttacaga	660	
Db	601	TTCTAGAAAATCTGATCTCGGAGAAATTCCTGCCCCAGACCCCTTCTGTTGTATACAGA	660	
Qy	661	acttcaacatctcgagactttaacgggaagtgtacattacaagtgttgtgaatccaacct	720	
Db	661	ACTTCAACATCTCGGACTTTACGGGAAGTGTGATATACAGTGGCTTTGAATCCAACCT	720	
Qy	721	tgtatgccttcgactccagctgcatagtgtccacacagaagtgtaacaagcttgtgtg	780	
Db	721	TTGATGCTTTCGACTGCCAGCTGTCATAGTTCACACAGAGAGGTGACAAACAGCTTGTG	780	
Qy	781	gaaacatctcttggagaaataagacacctagacagtggtattcttactagtgtaacccgtac	840	
Db	781	GAAACATCTCTGGGAATAAAGACCCCTAGACAGTGGATTTCTTTACTAGTCAACCGCTAC	840	
Qy	841	aaaaattcgtcaagatccttaaccaacctgggttctctcaaatcatgacaacagatacc	900	
Db	841	AAAAATTCGTGCAAGATCTTAACCAACCTGGTGTCTCTACANTCATGACAACGAGTACC	900	
Qy	901	ttcaatcaagatgactggtatactctgtcatcaaatagatagagaataaacctgaagact	960	
Db	901	TTCACTATCAAGATGACTGGTATATCTCTGTCATCAAGATAGAGATAAACCCTCAAGACT	960	
Qy	961	atatattgtactaccgtggcgaacagatgcttggatggatattggtgtgcattg	1020	
Db	961	ATATATTTGTACTACCGTGGCGGCAACGATGCTTGGGATGGATATGGTGGTCAGTTG	1020	
Qy	1021	tatacacgagaagttctgtattaccocaaatagcatattataccagaactcgaagaacagcaa	1080	

RESULT 2
US-08-747-574-2
; Sequence 2, Application US/08747574
; Patent No. 6015939
; GENERAL INFORMATION:
; APPLICANT: CALGENE, INC.
; TITLE OF INVENTION: PLANT VDE GENES AND
; TITLE OF INVENTION: METHODS RELATED THERETO
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.40 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION NUMBER:
; APPLICATION NUMBER: US/08/747,574
; FILING DATE: No. 6015939sember 7, 1996
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 60/023,502
; FILING DATE: August 6, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/006,315
; FILING DATE: No. 6015939sember 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: 119-2

	Query Match	3.0%	Score 46.4	DB 4	Length 3211
	Best Local Similarity	52.5%	Pred. No. 0.0012		
	Matches 126	Conservative 0	Mismatches 111	Indels 3	Gaps 1
Qy	1155	agaattgagaagacagtggagaaggtgaaaggataatcgtaaaagaggttgaacagata	1214		
Db	2392	AGTGATGAAGAGGAGGAGGAAGAGGAGGAAGGAAGGAAGGAAGGAAGGAAGAG	2451		
Qy	1215	gaagaagaaggttagagaaggaagtgaggaaggttcggttagactgagatgacctgtgtccag	1274		
Db	2452	GAAGNAGGAAGAGGAGAGACGAAAGGAGGGAAGACTTTTGAAGGAAGA	2508		
Qy	1275	agattggctgaaggtatttaatgaactgaagcaagcagcgaggagaatttcgtgagagagtta	1334		
Db	2509	GATGAAGAGGGAATATTTTTGAAGAGCAAGCAAGGAGGAGGAAGAGACTTTGAGCAAGAA	2568		

	Best Local Similarity	25.98;	Pred. No.	0.0064;	
	Matches	136;	Conservative	86;	Mismatches 308; Indels 0; Gaps 0;
QY	756	acagaagtgccacaacgaagtgttggaaacatctctggagaataaaagacccttagacagt	815		
Dd	226	ACNGTGCNNAARAARAARYTYNMGNAAYAAARACNAARGAARAGGNNAARAAYGARAAYSN	285		
QY	816	ggattcttaactagtcagcgcgtcaaaaaacctgccaagatctctaaccaacctggtgtt	875		
Dd	286	GARTGYCCNAARAARGARCCNAARCAACNAARCAACNAARCTGTGCGNGAYAAARAACNCMN	345		
QY	876	ctctacaatcatgacaacgagctacctcacatcaagatgactgttatctctatcatca	935		
Dd	346	YTNAATHATMGNAAYTNWSNTTYAARTGYWSNGARGAYGAYYYNAARACMGTWTTTCN	405		
QY	936	aagatagagaataaacctggaagactatatatttgtatactacogtggcgcaaacgatgct	995		
Dd	406	CARTTYGGNGCNGTYTNGARGTNAAYATHCCNMGNARCCNGAYGGNAARATGMNGGN	465		
QY	996	tgggatggatatggtggtgcagttgtatacacgagaagttctgtattaccaaatagcatt	1055		
Dd	466	TTYGNTTYTNCARTTYAARAAYTYNYTGARCNCNGNNAARCNCTYNAARGGNATGAAY	525		
QY	1056	ataccagaactcgaaaagcagcaaaaagcataggcagagacttcagcacattcattaga	1115		
Dd	526	ATGAARGARATHAARGGNMGNMACNGTNGCNGTNGAYTGGCGNGTNGCNAARGAYAARTAY	585		
QY	1116	acgataacacatggtgctcgtgaacctgcgtcgtggagagaattggagaagacagtggaa	1175		
Dd	586	AARGAYACNCARWSNGTYNSGCCNATHGGNGARGARAARWSNCAYGARWSNAARCAVCAR	645		
QY	1176	gaagtgtaaaaggataactcgttaaagaggttgaaagatagaagaagaggttagaagaa	1235		
Dd	646	GARWSNCTNAARAARAARGGNMGNARGARGARGAYATGGARGARGARGAAYGAYGAY	705		
QY	1236	gtggagaaggtcgtgtagcagctgagatgacctgtgtccagagattggtcgaag	1287		
Dd	706	GAYGAYGAYGAYGAYGARGARGAYGGNGTNTTYGAYGARGAYGARG	757		

RESULT 12
US-08-676-974-2
Sequence 2, Application US/08676974
Patent No. 5770422
GENERAL INFORMATION:
APPLICANT: COLLINS, KATHLEEN
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,974
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cdna
US-09-098-487-2

Query Match 2.8%; Score 43.6; DB 2; Length 2277;
Best Local Similarity 25.9%; Pred. No. 0.0064;
Matches 138; Conservative 86; Mismatches 308; Indels 0; Gaps 0;
QY 756 acagaaggtgacacagctgttggaacatctcttgagaataaagacccttagacagt 815
DB 226 ACNGTNCNAAARAAARTYNWGNAAACNAAARGAARAGNAAAYGARAAYWSN 285
QY 816 ggattcttactaggtcagccgtacaaaattcgtcgaagatctcaccacacctgggtt 875
DB 286 GARTGYCCNAAARAGCCNAAARGCNAARAGCNAARAGTNGCNGAYAAARAGCNMGN 345
QY 876 cctacaatcatgacacagtagtacctcactatcatcagagactggttatctcgtcatca 935
DB 346 YTNATHMGNAAAYTNWNTTYAARTGWSNGARGAGYAYTNAARACNGTNTTYGCGN 405
QY 936 aagatagagaataaacctgaaactatatattgttactacctggtggcgaaacgatgct 995
DB 406 CARTYGGCNGTNTYNGARGTNAAYATHCNMGNAARCCNGAYGNGNAARATGMNGGN 465
QY 996 tgggatgatatggtgagctgttatcacagagaagttctgtattaccacaaatagcatt 1055
DB 466 TTYGNTTYGTCARTTYAARAAYTYNTYNGRCNGGNNAARCNCTYNAARAGGNATGAAY 525
QY 1056 ataccagaactcgaagcagcaaaagcatagcagagactcagcattcattcattaga 1115
DB 526 ATGAARGARATHAARGGNMGNACNTGNCNGTNGAYTGGCNGTNGCNAAARGAAYARTAY 595
QY 1116 acgataacacatggtcctgaacctcgtcgtgagagaattgagaagacagtggaa 1175
DB 586 AARGAYACNARWSNGTWSNGCNATGGNGARGARAARWSNCAYGARWSNAAARCAVCAR 645
QY 1176 gaaggtgaaggaataatcgtaaaagaggttgagagataagaagaggtgagaagaa 1235
DB 646 GARWSNCTNAARAARAARGGNMGNARGARGARGAYATGGARGARGARGAAYGAYGAY 705
QY 1236 gtgagaaggtcgttaggactgagatgacctgttccagagagattgctgag 1287
DB 706 GAYGAYGAYGAYGAYGARGARGAYGNGTNTTYGAYGARGAYGARG 757

RESULT 14
US-08-728-323A-1
Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 2.8%; Score 43.2; DB 2; Length 3489;
Best Local Similarity 47.7%; Pred. No. 0.01;
Matches 126; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
QY 1198 aagaaggtgaagatagaagaagaggttagaagaaggaagtgagaggtcggtagactg 1257
DB 1094 ACAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATG 1153
QY 1258 agatgacctgttccagagagattggtcgaagatttaataaactgaacgaacagaggaga 1317
DB 1154 ACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1213
QY 1318 attcgtgagagagagtaagtaagaagaagagatggagtttttggatgagatcctaaatggaa 1377
DB 1214 ATGACGATGATGAGCAATGAGGAGGATGACGAGGAGGATGACGAGGAGGAGGAGGAGGAGG 1273
QY 1378 caagtgaagttgaaaaattgttgggaaagctttgccaatcaggaaggtccaggtagaac 1437
DB 1274 ACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1333
QY 1438 aagaacccaccattgtgtacaaac 1461
DB 1334 AGGAGCCCAACAGCAGGAGGCCAC 1357

RESULT 15
US-08-770-379-20/c
Sequence 20, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20

Search completed: November 6, 2001, 02:52:34
Job time: 11033 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 6, 2001, 00:54:09 ; Search time 3841.92 Seconds
(without alignments)
3826.000 Million cell updates/sec

Title: US-09-075-375A-5
Perfect score: 1555
Sequence: 1 ccacgcgccgctgtgtgt.....ggataaaaaaaaaaaaaa 1555

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST :
1: gb_est1 : *
2: gb_est2 : *
3: gb_est3 : *
4: gb_est4 : *
5: gb_est5 : *
6: gb_est6 : *
7: gb_est7 : *
8: gb_est8 : *
9: gb_est9 : *
10: gb_est10 : *
11: gb_est11 : *
12: gb_est12 : *
13: gb_est13 : *
14: gb_est14 : *
15: gb_est15 : *
16: gb_est16 : *
17: gb_est17 : *
18: gb_est18 : *
19: gb_est19 : *
20: gb_est20 : *
21: gb_est21 : *
22: gb_est22 : *
23: gb_est23 : *
24: gb_est24 : *
25: gb_est25 : *
26: gb_est26 : *
27: gb_est27 : *
28: gb_est28 : *
29: gb_est29 : *
30: gb_est30 : *
31: gb_est31 : *
32: gb_est32 : *
33: gb_est33 : *
34: gb_est34 : *
35: gb_est35 : *
36: gb_est36 : *
37: gb_est37 : *
38: gb_est38 : *
39: gb_est39 : *
40: gb_est40 : *
41: gb_est41 : *
42: gb_est42 : *
43: gb_est43 : *
44: gb_est44 : *
45: gb_est45 : *
46: gb_est46 : *
47: gb_est47 : *
48: gb_est48 : *
49: gb_est49 : *
50: gb_est50 : *
51: gb_est51 : *
52: gb_est52 : *
53: gb_est53 : *
54: gb_est54 : *
55: gb_est55 : *
56: gb_est56 : *
57: gb_est57 : *
58: gb_est58 : *
59: gb_est59 : *
60: gb_est60 : *
61: gb_est61 : *
62: gb_est62 : *
63: gb_est63 : *
64: gb_est64 : *
65: gb_est65 : *
66: gb_est66 : *
67: gb_est67 : *
68: gb_est68 : *
69: gb_est69 : *
70: gb_est70 : *
71: gb_est71 : *
72: gb_est72 : *
73: gb_est73 : *
74: gb_est74 : *
75: gb_est75 : *
76: gb_est76 : *
77: gb_est77 : *
78: gb_est78 : *
79: gb_est79 : *
80: gb_est80 : *
81: gb_est81 : *
82: gb_est82 : *
83: gb_est83 : *
84: gb_est84 : *
85: gb_est85 : *
86: gb_est86 : *
87: gb_est87 : *
88: gb_est88 : *
89: gb_est89 : *
90: gb_est90 : *
91: gb_est91 : *
92: gb_est92 : *
93: gb_est93 : *
94: gb_est94 : *
95: gb_est95 : *
96: gb_est96 : *
97: gb_est97 : *
98: gb_est98 : *
99: gb_est99 : *
100: gb_est100 : *
101: gb_est101 : *
102: gb_est102 : *
103: gb_est103 : *
104: gb_est104 : *
105: gb_est105 : *
106: gb_est106 : *
107: gb_est107 : *
108: gb_est108 : *
109: gb_est109 : *
110: gb_est110 : *
111: gb_est111 : *
112: gb_est112 : *
113: gb_est113 : *
114: gb_est114 : *
115: gb_est115 : *
116: gb_est116 : *

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			Description		
		Match	Length	ID			
1	463.6	29.8	483	104	AI994713	AI994713	701499353
C 2	400	25.7	400	30	AV440942	AV440942	AV440942
C 3	381.6	24.5	424	30	AV442545	AV442545	AV442545
C 4	330.8	21.3	335	30	AV521985	AV521985	AV521985
C 5	322.8	20.8	438	159	N37612	N37612	18839 Lambd
C 6	263.8	17.0	515	144	BF069291	BF069291	st45b12.y
C 7	256	16.5	599	256	B28051	B28051	T2G7nFB TA
C 8	244.6	15.7	1147	256	B08217	B08217	T27M8-T7.1
C 9	243	15.6	1147	256	BI3003	BI3003	T27G7-T7.1
C 10	237.4	15.3	935	256	BI2210	BI2210	T27G7-T7 TA
C 11	235.6	15.2	427	30	AV410579	AV410579	AV410579
C 12	231.2	14.9	985	256	BI3009	BI3009	T27M8-T7 TA
C 13	174.6	11.2	549	141	BE918838	BE918838	FW1_2_F01
C 14	164.8	10.6	602	23	A1668224	A1668224	605018D09
C 15	151	9.7	411	238	AZ126934	AZ126934	OSJNMB007
C 16	151	9.7	812	238	AZ126249	AZ126249	OSJNMB007
C 17	149.4	9.6	871	234	AQ858812	AQ858812	nbe50002N
C 18	149	9.6	248	30	AV410854	AV410854	AV410854
C 19	138.4	8.9	319	164	BE191198	BE191198	sn93d05.y
C 20	130.4	8.4	647	115	AW398615	AW398615	EST309115
C 21	118	7.6	418	137	BE959698	BE959698	P11_55_E1
C 22	116.2	7.5	724	174	BG123331	BG123331	EST468977
C 23	104.8	6.7	584	141	BE918982	BE918982	FW1_2_F01
C 24	94.6	6.1	383	167	BE440946	BE440946	sp30a05.y
C 25	88.6	5.7	483	151	BF623692	BF623692	HVSMea000
C 26	79	5.1	997	219	CNS005TE	CNS005TE	Drosophil
C 27	75.2	4.8	112	174	BG157169	BG157169	sab33d04.
C 28	72.2	4.6	584	20	A483033	A483033	EST242356
C 29	68.8	4.4	572	122	AW929268	AW929268	EST338056
C 30	64	4.1	1101	219	CNS0039G	CNS0039G	Drosophil
C 31	63.8	4.1	1101	219	CNS0181N	CNS0181N	Drosophil
C 32	62.8	4.0	987	219	CNS00418	CNS00418	Drosophil
C 33	57.2	3.7	1101	219	CNS00AWD	CNS00AWD	Drosophil
C 34	56.6	3.6	669	151	BF641377	BF641377	NF061G021
C 35	56	3.6	1101	219	CNS017RP	CNS017RP	Drosophil
C 36	54.6	3.5	1101	219	CNS00GCG	CNS00GCG	Drosophil
C 37	53.8	3.5	863	219	CNS0169F	CNS0169F	Drosophil
C 38	53.8	3.5	938	219	CNS006TJ	CNS006TJ	Drosophil
C 39	53.6	3.4	662	249	AZ765193	AZ765193	1M0562E01
C 40	53.4	3.4	1204	219	CNS016E2	CNS016E2	Drosophil
C 41	53.2	3.4	1101	219	CNS0182P	CNS0182P	Drosophil
C 42	53	3.4	1100	219	CNS00EEY	CNS00EEY	Drosophil
C 43	52.8	3.4	959	219	CNS00655	CNS00655	Drosophil
C 44	52.4	3.4	1101	219	CNS0106X	CNS0106X	Drosophil
C 45	52.2	3.4	861	256	R20015	R20015	FW12A20-T7.1

ALIGNMENTS

[illegible]

REFERENCE Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 400)
 AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
 JOURNAL DNA Res. 7, 175-180 (2000)
 MEDLINE 20363093
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 FEATURES Location/Qualifiers
 source
 1..400
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="AP216c07_f"
 /clone_lib="Arabidopsis thaliana above-ground organ two to six-week old"
 /tissue_type="aboveground organs"
 /dev_stage="two to six-week old"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 81 a 120 c 48 g 151 t
 ORIGIN

Query Match 25.7%; Score 400; DB 30; Length 400;
 Best Local Similarity 100.0%; Pred. No. 5.7e-96;
 Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1065 CTCGAAAAGCAGCAAAAGCATAGGACAGCTTCAGACATTCATTAGACGGATAAC 1124
 DB 400 CTCGAAAAGCAGCAAAAGCATAGGACAGCTTCAGACATTCATTAGACGGATAAC 341
 QY 1125 ACATGTGCTCAGCAAAAGCATAGGACAGCTTCAGACATTCATTAGACGGATAAC 1184
 DB 340 ACATGTGCTCAGCAAAAGCATAGGACAGCTTCAGACATTCATTAGACGGATAAC 281
 QY 1185 AGGATAATCGTAAAGAGGTTGAAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAG 1244
 DB 280 AGGATAATCGTAAAGAGGTTGAAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAG 221
 QY 1245 GTCGGTAGGACTGAGATGACCTTGTCCAGAGATTCGCTGAAGGATTTAATGAAGTGAAG 1304
 DB 220 GTCGGTAGGACTGAGATGACCTTGTCCAGAGATTCGCTGAAGGATTTAATGAAGTGAAG 161
 QY 1305 CAAGACGAGAGAGATTCGAGAGAGATTCGAGAGAGATTCGAGAGAGAGAGAGAGAGAG 1364
 DB 160 CAAGACGAGAGAGATTCGAGAGAGATTCGAGAGAGATTCGAGAGAGAGAGAGAGAGAG 101
 QY 1365 ATCAAAATGAAAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1424
 DB 100 ATCAAAATGAAAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 41
 QY 1425 GTCAGGTAG 1464
 DB 40 GTCAGGTAGAT 1

RESULT 3
 AV442545 424 bp mRNA EST 14-NOV-2000
 LOCUS AV442545 Arabidopsis thaliana above-ground organ two to six-week
 DEFINITION old Arabidopsis thaliana cDNA clone AP216c07_r 5', mRNA sequence.
 ACCESSION AV442545
 VERSION AV442545.1 GI:7612957
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 424)
 AUTHORS Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
 JOURNAL DNA Res. 7, 175-180 (2000)
 MEDLINE 20363093
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 FEATURES Location/Qualifiers
 source
 1..424
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="AP216c07_r"
 /clone_lib="Arabidopsis thaliana above-ground organ two to six-week old"
 /tissue_type="aboveground organs"
 /dev_stage="two to six-week old"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 97 a 86 c 99 g 142 t
 ORIGIN

Query Match 24.5%; Score 381.6; DB 30; Length 424;
 Best Local Similarity 99.0%; Pred. No. 4.7e-91;
 Matches 384; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 7 GTCGGTGTGTTGGGAGAGATTAGATAGTGTGAAGATGGCAGTAGCTACACATTGTT 66
 DB 37 GTATTGCTGTGTTGGGAGAGATTAGATAGTGTGAAGATGGCAGTAGCTACACATTGTT 96
 QY 67 TCATTACCTTGTGATGACCGTATTCGATTTTCTCAAGTGTGATGATGATGATGATGATG 126
 DB 97 TCATTACCTTGTGATGACCGTATTCGATTTTCTCAAGTGTGATGATGATGATGATGATG 156
 QY 127 TGGCATTACAGAAAGAGGATCAATGGCATTCTTCTCAAGTGTGATGATGATGATGATGATG 186
 DB 157 TGGCATTACAGAAAGAGGATCAATGGCATTCTTCTCAAGTGTGATGATGATGATGATGATG 216
 QY 187 AAAGTGTGATCTCAGAACAACTGGTGGAGATCTCAGCGTCTTATCTGCAATTCAGGT 246
 DB 217 AAAGTGTGATCTCAGAACAACTGGTGGAGATCTCAGCGTCTTATCTGCAATTCAGGT 276
 QY 247 CAGGATCTCTAAGGGGATATTGACATTTGGCCATTACCATCAAGAAATGAGCTGAAG 306
 DB 277 CAGGATCTCTAAGGGGATATTGACATTTGGCCATTACCATCAAGAAATGAGCTGAAG 336
 QY 307 AGCTGACCGCTCCGCTGTGCTAAAGTCTGGGTGTTTGGTTCGCTTCCTATTG 366
 DB 337 AGCTGACCGCTCCGCTGTGCTAAAGTCTGGGTGTTTGGTTCGCTTCCTATTG 396
 QY 367 TCCATCTGCGAGATTCAGTTGATGCACT 394
 DB 397 TCCATCTGCGAGATTCAGTTGATGCACT 424

RESULT 4
 AV521985/c 335 bp mRNA EST 07-SEP-2000
 LOCUS AV521985 Arabidopsis thaliana aboveground organs two to six-week
 DEFINITION old Arabidopsis thaliana cDNA clone AP270c09F 3', mRNA sequence.
 ACCESSION AV521985
 VERSION AV521985.1 GI:8681512
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1 (bases 1 to 335)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries

JOURNAL

MEDLINE

COMMENT

Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 252-0812, Japan
Email: asamizuekazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers

FEATURES

source

1. .335
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AP270c09p"
/clone_lib="Arabidopsis thaliana aboveground organs two to six-week old"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
XhoI

BASE COUNT

85 a 90 c 39 g 121 t

Query Match

Best Local Similarity 21.3%; Score 330.8; DB 30; Length 335;

Matches

332; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1209 gagatagaagaagggtagagaaggaagtgagaagtgctgtaggactgagatgaccttg 1268

|||||

Db 335 GAGATAGAAGAAGGTAGAGAAGGAAGTGGAGAGTCTGGTAGGACTGAGATGACCTTG 276

Qy 1269 ttccagagattgctggaagatttaataaactgaagcaagcagagaggaatttcgtgaga 1328

|||||

Db 275 TTCAGAGATTGGCTGAAGGATTTAATGAACACTGAAGCAAGCAGGAGGAAATTCGTGAGA 216

Qy 1329 gaggtaagtaagaagagatgaggttttttgatgagatcaaaatgaagcaagtgaggtt 1388

|||||

Db 215 GAGTTAGTAAGAAGAGATGGAGTCTTTGGATGAGATCAAAATGGAAGCAAGTGAGGTT 156

Qy 1389 gaaaaattgtttgggaagcgtttgccaatcaggaaggtcaggttagaacaagaccacca 1448

|||||

Db 155 GAAAAATTGTTGGGAAGCTTTGCCAATCAGGAAGTCAAGGTAGAAACAAGAACCCCA 96

Qy 1449 ttgtgtacaacataattatacaactgctgctcggttcataataagtaatttttga 1508

|||||

Db 95 TTGTTGTACAAACTATATTATACATCTGTGCGGTTCATATCAAGTAATATTTTGTGA 36

Qy 1509 cacagtcacatcatcattccatacaaatggataaa 1542

|||||

Db 35 CACAGTCATCATCTATCCATAACATTTGGATACA 2

RESULT

5

LOCUS

N37612

DEFINITION

18839 Lambda-PRL2 Arabidopsis thaliana cDNA clone 207C23T7, mRNA

sequence.

ACCESSION

N37612

VERSION

N37612.1

KEYWORDS

EST.

SOURCE

thale cress.

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

1 (bases 1 to 438)

TITLE

Genes galore: a summary of methods for accessing results from

large-scale partial sequencing of anonymous Arabidopsis cDNA clones

Plant Physiol. 106, 1241-1255 (1994)

95148729

JOURNAL

MEDLINE

COMMENT

Contact: Thomas Newman

MSU-DOE Plant Research Laboratory

Michigan State University

MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.

Lansing, MI

Tel.: 517-353-0854

Fax: 517-353-9168

Email: 22313tcn@bm.cl.msu.edu

Seq primer: T7 dye primer.

Location/Qualifiers

1. 438

FEATURES

source

/organism="Arabidopsis thaliana"

/strain="var columbia"

/db_xref="taxon:3702"

/clone="207C23T7"

/clone_lib="Lambda-PRL2"

/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;

Lambda PRL2 is a cDNA library derived from equal

quantities of 4 pools of mRNA. The mRNA sources were 1) 7

day germinated etiolated seedlings; 2) tissue culture

grown roots; 3) staged plants half with 24 hour light

cycle, half on 16 hr light, 8 hour dark- rosettes; 4)

same plants as 3 but aerial tissue (stems, flowers and

siliques. The vector is BRL's lambda Zip-Lox. The cDNA

inserts were directionally cloned with Sal-Not arms using

oligo dt primed cDNA.

BASE COUNT 108 a 97 c 102 g 114 t 17 others

ORIGIN

Query Match

20.8%; Score 322.8; DB 159; Length 438;

Best Local Similarity

92.5%; Pred. No. 2.4e-75;

Matches

358; Conservative 0; Mismatches 24; Indels 5; Gaps 2;

Qy 426 agatagaaactcgaaaagtgcaattgccaaacctgctgtgcagccaatgctgcgtgcctt 485

|||||

Db 14 AGATAGAAACTCGAAAAGTGCAATGCCAACCTGCTGTGCAGCCAATGTCGCGTGCCTT 73

Qy 486 cagacctgcaataaacctgcctccagatgaaccgagtgccagattaaatgtggggtctgttt 545

|||||

Db 74 CAGACCTGCAATAACCGTCCAGATGAACCGAGTGCCAGATTAAATGTGGGATCTGTTT 133

Qy 546 gagaacagtggttgatgagtgagtgcaacgagtgctgtgtgcagagaaaaaagtgttctct 605

|||||

Db 134 GAGAACAGTGTGTTGTATGATGAGTTCAACGAGTGTGCTGTGTCGAGAAAAAGTGTCTCT 193

Qy 606 agaaaatctgactctcgagaatttctgccagacctcttcttctgttctgtacagaaactc 665

|||||

Db 194 AGAAATCTGATCTCGGAGAAATTTCTGCCCCAGACCTCTCTGTTCTGTGTACAGAACTTC 253

Qy 666 aacatctcggaactttaacggggaagtgttacattacaagtggcttgaatcccaacctttgat 725

|||||

Db 254 AACATCTCGGACTTTAACGGGAAGTGGTACATTACAAGTGGCTTCAATNCAACCTTTGAT 313

Qy 726 gactctgac-tgcagctgcatgagttccacacagaaagtgacaaacag-cttctgttg 780

|||||

Db 314 GCCTTCGNCCTGCCANCTGCATGGGTTTTCCNACAGGAAGGTGCAACAGGCTTTNTTTGG 373

Qy 781 gaaacatctcttgagaaataaagacc 807

|||||

Db 374 AACAAACCNTTGGGGNATAAAGCCCC 400

RESULT

6

BF069291

LOCUS BF069291 515 bp mRNA 17-OCT-2000
 DEFINITION st45b12.v1 Gm-cl067 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-cl067-1991 5' similar to FR:Q40593 Q40593 VIOLAXANTHIN

DE-EPOXIDASE PRECURSOR. ; mRNA sequence.

ACCESSION BF069291

VERSION

KEYWORDS

SOURCE

ORGANISM

soybean.

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

1 (bases 1 to 515)

Shoemaker, R., Keim, P., Vodkin, L., Erpelind, J., Corvelli, V., Khanna

A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,

Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers

, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk

, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann

, R., Waterston, R. and Wilson, R.

Public Soybean EST Project

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World

Parkway Circle St. Louis, Missouri 63134 For further information

call: (800) 420-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)

427-3324 or contact: clones@genomesystems.com or

info@genomesystems.com web site: www.genomesystems.com

High quality sequence stop: 421.

FEATURES

Location/Qualifiers

1..515

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl067-1991"

/clone_lib="Gm-cl067"

/tissue_type="Germinating shoot, 3 day old seedling, auxin

treatment"

/lab_host="DH10B"

/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:

XhoI; The cDNA library was constructed from mRNA isolated

from germinating shoots of 3 day old seedling for the

cultivar Williams 82. The seedlings were germinated in a

growth chamber using germination paper in a solution

containing 100ppm auxin. Complementary DNA was synthesized

from mRNA using a primer consisting of a poly(AT) sequence

with a XhoI restriction site. EcoRI adapters were ligated

to the blunt-ended cDNA fragments followed by XhoI

digestion. The cDNA fragments were directionally cloned

into the EcoRI-XhoI restriction site of the pBluescript

vector. The ligated cDNA fragments were transformed into

DH10B host cells (GibcoBRL). This library was constructed

in the laboratory of Dr. Randy Shoemaker."

186 a 74 c 136 g 119 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 17.0%; Score 263.8; DB 144; Length 515;

Matches 371; Conservative 0; Mismatches 142; Indels 12; Gaps 1;

QY 905 ctatcaagatgactggtatctctcatcaagaatagataaacctgaagactatat 964

|||||

Db 2 CTATCAAGATGACTGCTATATTTTCTCATCCCAAGATTGAGATAAGCCAGATGACTACGT 61

|||||

QY 965 atttgatactaccgtggcgaaacagctctggatgagatggtggtgagttgtata 1024

|||||

Db 62 ATTGTGTATACCCGAGGAGCAATGATGCATGGGATGTTATGAGGTGCTGTATATA 121

|||||

QY 1025 cagcagaagttctgtattaccacaatagcattaccagaaactcgaaaaagcagcaaaag 1084
 |||||

Db 122 CACAAGAGTGCAGTTTTCCTGAATCTATAGTCTCTGACTTTGAAAAGAGCAGTAAGAG 181
 |||||

QY 1085 catagcagagacttcagcacattcattagaaacggaatacacatcttggtctgaacctgc 1144
 |||||

Db 182 TGTGGGAAGAGACTTTAGCACGTTTCATCAGGACAGATAACACATGCGGGCCAGAGCCTTC 241
 |||||

QY 1145 gctcgtgagagagaatcgagaagacagtggaagaagtggaagaagataatcgtaaaaaggt 1204
 |||||

Db 242 CCTGGTGAAGAGCTGGAGAAAAGGTAGAGGAAGGAGAGACCACTGTGAAGGGAAGT 301
 |||||

QY 1205 tgaagacatgaagaagaggttagaagaagtggaagagtcggttagactgagatgac 1264
 |||||

Db 302 TGAACAGTTAGAGAAGAGAG-----GTGGAGAGAGTGGGGAAAACGGAGGCCAC 349
 |||||

QY 1265 ctgttccagagattggttgaaagatttaatagaactgaagcagagaggaatttcgt 1324
 |||||

Db 350 CTGTGTTTCAAAAATTTGCAGAAAAGGTTCAAAAGTATTCCAAAAAGATGAAGAGAAATTTCTT 409
 |||||

QY 1325 gacagatttaagtaagaagagagatgagtttttggtgagatcaaaatggaagcaagtga 1384
 |||||

Db 410 AGAAGGTTATCAAAACAGAAATGGAATTTCTTTATAGCTCAAAATGGAACCCAGTG 469
 |||||

QY 1385 ggttgaaaaattgtttgggaaagctttgccaatcaggaaaggtcag 1429
 |||||

Db 470 AATAAAAAAACTCTTTAGACGTCGGCTTCGGGTGAAGAAACTAAG 514
 |||||

RESULT 7

B28051/c

LOCUS

DEFINITION

B28051 599 bp DNA GSS 13-OCT-1997

T27G7TFB TAMU Arabidopsis thaliana genomic clone T27G7, DNA

sequence.

ACCESSION B28051

VERSION B28051.1

GI:2514017

KEYWORDS

GSS.

SOURCE

thale cress.

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 599)

REFERENCE

AUTHORS

Rounsley, S.D., Kelley, J.M., Field, C.E., Craven, M.B., Adams, M.D. and

Venter, J.C.

Use of a BAC End Sequence Database To Identify Minimal Overlaps for

Arabidopsis Genomic Sequencing

Unpublished (1997)

JOURNAL

COMMENT

Other_GSSs: T27G7TR

Contact: Steve Rounsley

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: rounsley@tigr.org

Seq primer: M13-21

Class: BAC ends

High quality sequence stop: 599.

FEATURES

Location/Qualifiers

1..599

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/clone="T27G7"

/sex="hermaphrodite"

/note="Vector: BclBACII; Site_1: HindIII; Site_2: HindIII

; Produced by Rod Wing"

BASE COUNT

ORIGIN

195 a 130 c 111 g 163 t

Query Match 16.5%; Score 256; DB 256; Length 599;
 Best Local Similarity 100.0%; Pred. No. 1.7e-57;
 Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 522 cagattaaatgtgggatctgtttgagaacagtggtttgatgagttcaacagtgct 581
 |||||||
 DB 256 CAGATTAAATGTGGGATCTGTTTGTGAGAACAGTGTGTTGATGAGTTCAACGAGTGTGCT 197
 |||||||

QY 582 gtgtcgagaaaaagtggttcttagaaaaatctgatctcgagagaatttctgtccccagac 641
 |||||||
 DB 196 GTGTGAGAAAAAGTGTGTTCTAGAAAATCTGATCTCGAGAAATTTCTGCCCCAGAC 137
 |||||||

QY 642 ccttctgtctgtacagaactccaacatctcgactttaacggaagtgtacattaca 701
 |||||||
 DB 136 CCTTCTGTTCTGTACAGAACTTCAACATCTCGGACTTTAACGGGAAGTGGTACATTACA 77
 |||||||

QY 702 agtggcttgaatccaaacctttgatgccttcgactgacgtgcagctgcattgattccacacagaa 761
 |||||||
 DB 76 AGTGGCTTGAATCAACACCTTTTGATGCCTTCGACTGCCAGCTGCATGATGATCCACACAGAA 17
 |||||||

QY 762 ggtgacaacaagcttg 777
 |||||||
 DB 16 GGTGACAACAAGCTTG 1
 |||||||

RESULT 8
 B08217/c
 LOCUS B08217 1147 bp DNA GSS 14-MAY-1997
 DEFINITION T27M8-T7.1 TAMU Arabidopsis thaliana genomic clone T27M8, DNA sequence.

ACCESSION B08217
 VERSION B08217.1 GI:2089494
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 1147)
 AUTHORS Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and Ecker, J.
 TITLE BAC End Sequences at ATGC
 JOURNAL Unpublished (1997)
 COMMENT Other GSSs: T27M8-Sp6.1, T27M8-Sp6, T27M8-T7
 Contact: Ecker J.
 Arabidopsis Thaliana Genome Center
 University of Pennsylvania
 Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104
 Tel: 215-898-9384
 Fax: 215-898-8780
 Email: jecker@atgenome.bio.upenn.edu
 Seq primer: T7
 Class: BAC ends
 High quality sequence start: 110
 High quality sequence stop: 594.

FEATURES
 source
 1. .1147
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="T27M8"
 /clone_lib="TAMU"
 /sex="hermaphrodite"
 /note="Vector: BelobACII; Site_1: HindIII; Site_2: HindIII
 ; Produced by Rod Wing"
 BASE COUNT 204 a 366 c 195 g 357 t 25 others
 ORIGIN

Query Match 15.7%; Score 244.6; DB 256; Length 1147;
 Best Local Similarity 98.0%; Pred. No. 2.2e-54;
 Matches 247; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 522 cagattaaatgtgggatctgtttgagaacagtggtttgatgagttcaacagtgct 581
 |||||||
 DB 320 CAGATTAAATGTGGGATCTGTTTGTGAGAACAGTGTGTTGATGAGTTCAACGAGTGTGCT 261
 |||||||

QY 582 gtgtcgagaaaaagtggttcttagaaaaatctgatctcgagagaatttctgtccccagac 641
 |||||||
 DB 260 GTCTCGAGAAAAAGTGTGTTCTAGAAAATCTGATCTCGGAGAAATTTCTGCCCCAGAC 201
 |||||||

QY 642 ccttctgttctgtacagaactccaacatctcgactttaacggaagtgtacattaca 701
 |||||||
 DB 200 CCTTCTGTTCTGTACAGAACTTCAACATCTCGGACTTTAACGGGAAGTGGTACATTACA 141
 |||||||

QY 702 agtggcttgaatccaaacctttgatgccttcgactgacgtgcagctgcattgattccacacagaa 761
 |||||||
 DB 140 AGTGGCTTGAATCAACACCTTTTGCTTCCGANTCCAGCCCCATGAGTTCACACAGAA 81
 |||||||

QY 762 ggtgacaacaag 773
 |||||||
 DB 80 GGTGACAACAAG 69
 |||||||

RESULT 9
 B13003/c
 LOCUS B13003 1147 bp DNA GSS 14-MAY-1997
 DEFINITION T27G7-T7.1 TAMU Arabidopsis thaliana genomic clone T27G7, DNA sequence.

ACCESSION B13003
 VERSION B13003.1 GI:2094135
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 1147)
 AUTHORS Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and Ecker, J.
 TITLE BAC End Sequences at ATGC
 JOURNAL Unpublished (1997)
 COMMENT Other GSSs: T27G7-Sp6, T27G7-T7, T27G7-Sp6.1
 Contact: Ecker J.
 Arabidopsis Thaliana Genome Center
 University of Pennsylvania
 Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104
 Tel: 215-898-9384
 Fax: 215-898-8780
 Email: jecker@atgenome.bio.upenn.edu
 Seq primer: T7
 Class: BAC ends
 High quality sequence start: 70
 High quality sequence stop: 564.

FEATURES
 source
 1. .1147
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="T27G7"
 /clone_lib="TAMU"
 /sex="hermaphrodite"
 /note="Vector: BelobACII; Site_1: HindIII; Site_2: HindIII
 ; Produced by Rod Wing"
 BASE COUNT 249 a 372 c 198 g 319 t 9 others
 ORIGIN

Query Match 15.6%; Score 243; DB 256; Length 1147;
 Best Local Similarity 96.1%; Pred. No. 6e-54;
 Matches 249; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 522 cagattaaatgtgggatctgtttgagaacagtggtttgatgagttcaacagtgct 581
 |||||||


```

* QY 536 gqatctgtttgagaaacagtggtttgtgagtgatccaacagtggtgtgtgagaaaaa 595
  Db 242 GGACCTGTTTGAACACAGTGTGGTTGATCAATTTAATGAGTGTGAGTCTCCGGAAGAA 301
  QY 596 gtgtgtcttcagaaaaatctgatctgcctggagaatttctctgccagacccttctgttctt 655
  Db 302 ATGTGTACCTAAGAAATCTGACGTGGAGAGTTTCTCTGCTCCAAATCCTGATGCTCTGT 361
  QY 656 acagaaacttcaacatctcgagatttaacaggggaagtgggtacattacaagtggctgaatcc 715
  Db 362 GAAGAGCTTCAATATATGCGAGATTTCACTGCGCAAGTGGTTCATCACTAGTGGCTTAAATCC 421
  QY 716 aacctt 721
  Db 422 TACCTT 427

RESULT 12
B13009/C B13009 985 bp DNA GSS 14-MAY-1997
LOCUS T27M8-T7 RAMU Arabidopsis thaliana genomic clone T27M8, DNA
DEFINITION sequence.
ACCESSION B13009
VERSION B13009.1 GI:2094141
KEYWORDS GSS,
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 985)
AUTHORS Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
Ecker,J.
TITLE BAC End Sequences at ATGC
JOURNAL Unpublished (1997)
COMMENT Other GSSs: T27M8-T7.1, T27M8-Sp6.1, T27M8-Sp6
Contact: Ecker J.
Arabisopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 350
High quality sequence stop: 477.

FEATURES
source
1..985
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T27M8"
/clone_lib="TAMU"
/sex="hermaphrodite"
/note="Vector: BelOBACII; Site_1: HindIII; Site_2: HindIII
; Produced by Rod Wing"
BASE COUNT 243 a 275 c 120 g 272 t 75 others
ORIGIN
14.9%; Score 231.2; DB 256; Length 985;
Best Local Similarity 95.9%; Pred. No. 8.2e-51;
Matches 233; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 535 gggatctgtttgagaaacagtggtttgtgagtgatccaacagtggtgtgtgagaaaaa 594
Db 341 GGGATCTGTTTGAACACAGTGTGGTTGATGAGTTCACAGAGTGTGCTGTGCGAGAAAAA 282
QY 595 agtgtgtcttcagaaaaatctgatctcgagaaatttctgccagacccttctgttcttg 654
Db 281 NGTGTGTCNTAGANNATCTGATCTCGAGAAATTCNTGCCCGACACCTTCTCTGTTCTTG 222

```

```

QY 655 tacagaacttcaacatctcgagatttaacaggggaagtgggtacattacaagtgttgaatc 714
  Db 221 TACAGCATTTCAACATCTCGACTTTTACCGGAAGTGGTACATTACAAAGTGGCTTGAATC 162
  QY 715 caaccttgccttcgagctgcagctgcagctgcagctgcagctgcagctgcagctgcagctgc 774
  Db 161 CAACCTTTGATGCTTGCAGCTGCAGCTGCAGCTGCAGCTGCAGCTGCAGCTGCAGCTGCAG 102
  QY 775 ttg 777
  Db 101 TTG 99

```

RESULT 13

```

BE918838 BE918838 549 bp mRNA EST 29-SEP-2000
LOCUS FM1_2_F01.g1_A003 Floral-Induced Meristem 1 (FM1) Sorghum
DEFINITION propinquum cDNA, mRNA sequence.
ACCESSION BE918838
VERSION BE918838.1 GI:10422122
KEYWORDS EST.
SOURCE Sorghum propinquum.
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 549)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
,L.H.
TITLE An EST database from Sorghum: floral-induced meristems
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix
High quality sequence start: 89
High quality sequence stop: 506
POLYA-No. Location/Qualifiers
1..549
/organism="Sorghum propinquum"
/db_xref="taxon:132711"
/clone_lib="Floral-Induced Meristem 1 (FM1)"
/note="Organ: Floral-Induced meristems; Vector:
pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; mature plants were placed in a growth chamber for
15 days with 16 hr darkness and 8 hr light (flowering is
induced by short-day conditions); 16 days after being
returned to the greenhouse under natural long days during
late April/early May, meristems were harvested. The
library was made from poly-A RNA in the cloning vector
lambda ZAP II. Clones to be sequenced were prepared by
mass excision."
BASE COUNT 165 a 97 c 155 g 132 t
ORIGIN
Query Match 11.2%; Score 174.6; DB 141; Length 549;
Best Local Similarity 66.6%; Pred. No. 9.1e-36;
Matches 271; Conservative 0; Mismatches 124; Indels 12; Gaps 1;

QY 1057 taccagaactcgaagaaacagcaaaaagcagagacagacagacattcagcattcattagaa 1116
  Db 1 TACCTGAGCTAGAAAGAGCTGCAGAAAGCATAGTCTCGGACTCTCTCGACGTTTCATCAGA 60

```


times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, the Nipponbare variety using *EcoRI* as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu).

```

BASE COUNT      108 a      91 c      100 g      112 t
ORIGIN

Query Match      9.7%; Score 151; DB 238; Length 411;
Best Local Similarity 75.7%; Pred. No. 1.7e-29;
Matches 187; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 523 agattaaatgtggggtctgttttgagaacagtggtgtgatgagttcaacgagtggtgctg 582
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 270 AGATCAAAATGTGGAGATCTGTTTGAGACACACTGTGTCGATGAGTTCAACGAGTGTGCTG 211

QY 583 tgtcgagaaaaaagtgtgttctctagaaaaatctgatctcgagaaatttcctgccccagacc 642
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210 TTTGCGGCAAGAAATGCGTCCCAACAAAAGTCCGACGTTGGCGAGTTCCCGAGTCCCTGATC 151

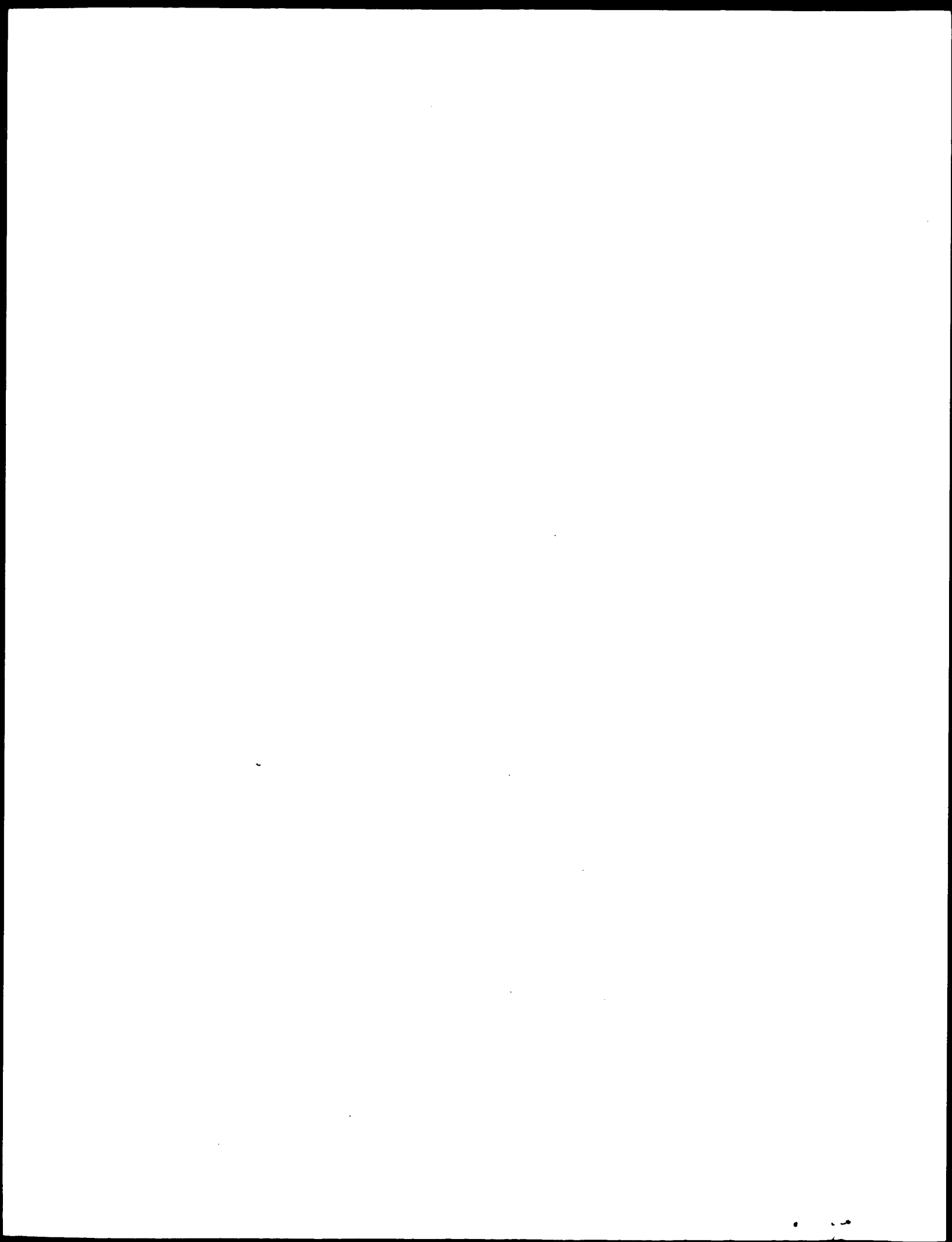
QY 643 ctctctgtctgttacagaacttcaacatctcggaactttaacgggaagtgtgtacattacaa 702
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150 CATCCGCCCTTGTCAAGAACTTCAACATGGCTGATTTCAACGGCAAGTGTATATTTCAA 91

QY 703 gtggttgaaatccaacotttgatgccttcgactgcagctgcatgaggttccacacagaag 762
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 90 GTGGCCTCAATCCCACTTTTCGACACATTCGATTGCCAACTTCACGAGTTCCTGTGCGAGG 31

QY 763 gtgacaa 769
      |||||
Db 30 GAGACAA 24

```

Search completed: November 6, 2001, 00:54:15
Job time: 4104 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 04:59:38 ; Search time 93.61 Seconds
(without alignments)
914.444 Million cell updates/sec

Title: US-09-075-375A-6
Perfect score: 7495
Sequence: 1 MALSHTVFLCKEALNLYA.....MEASEVKLFGKALPIRKVR 1412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2554	34.1	478	18	AAW09875 Tobacco violaxanthin
2	2498	33.3	473	18	AAW09874 Romaine lettuce vi
3	2417	32.2	462	18	AAW09876 Arabidopsis violax
4	212.5	2.8	525	21	AAG49996 Arabidopsis thalia
5	210	2.8	522	21	AAG49997 Arabidopsis thalia
6	209	2.8	522	21	AAG06331 Arabidopsis thalia
7	209	2.8	526	21	AAG06330 Arabidopsis thalia
8	207	2.8	428	21	AAG49998 Arabidopsis thalia
9	206	2.7	428	21	AAG06332 Arabidopsis thalia
10	151	2.0	1055	21	AAV44787 Arabidopsis thalia
11	147.5	2.0	1786	18	AAW24790 P. falciparum live

12	147	2.0	2482	16	AAW72826 Human mitosin. HO
13	147	2.0	2482	19	AAW23996 Human mitosin amin
14	143.5	1.9	1558	21	AAW18324 Plasmodium falcipa
15	135.5	1.8	1312	18	AAW22775 Human RAD50. Homo
16	135.5	1.8	1312	19	AAW71295 Human homologue of
17	135	1.8	968	19	AAW85011 p85alpha-green flo
18	134	1.8	1639	19	AAW54145 P. falciparum synt
19	133	1.8	1979	21	AAW54145 Plasmodium falcipa
20	132	1.8	776	15	AAW18171 Lethal factor of B
21	129.5	1.7	2954	20	AAW60178 Amino acid sequenc
22	129	1.7	993	17	AAW95268 Pre-nisin modifica
23	129	1.7	993	20	Nisin B of Lactoba
24	129	1.7	1654	6	AAW50777 Sequence of the p1
25	128.5	1.7	1521	21	AAW39235 Arabidopsis thalia
26	128.5	1.7	1528	21	AAW39234 Arabidopsis thalia
27	128.5	1.7	1562	21	AAW39233 Arabidopsis thalia
28	128.5	1.7	1703	21	AAW36714 Arabidopsis thalia
29	128.5	1.7	1710	21	AAW36713 Arabidopsis thalia
30	128.5	1.7	1744	21	AAW36712 Arabidopsis thalia
31	124	1.7	1498	16	AAW77084 Rat sulphonylurea
32	124	1.7	1582	16	AAW77087 Rat sulphonylurea
33	121.5	1.6	1132	17	AAW97866 Chicken leucocyt
34	121.5	1.6	1369	20	AAW24788 Human secreted pro
35	121	1.6	1411	17	AAW02258 Nucleolar/endosoma
36	119.5	1.6	1099	19	AAW40538 Mutant c-beta prot
37	119.5	1.6	1604	16	AAW70105 TNF-R-EBA 175 fusi
38	119.5	1.6	1847	21	AAW52002 M. jannaschii MJ14
39	119.5	1.6	1847	21	AAW51631 M. jannaschii MJ14
40	119.5	1.6	1881	21	AAW44506 Streptococcus pneu
41	119	1.6	905	18	AAW31186 Human p160 polyep
42	119	1.6	1135	18	AAW31185 Human p160 polyep
43	118.5	1.6	2273	17	AAW98811 Erysiphe graminis
44	117.5	1.6	1230	21	AAW49306 Potato starch synt
45	117.5	1.6	3542	22	AAW62142 P. falciparum FCR3

ALIGNMENTS

RESULT 1
AAW09875
ID AAW09875 standard; Protein; 478 AA.
XX
XX
AC AAW09875;
XX
XX
DT 28-JUL-1997 (first entry)
XX
XX
DE Tobacco violaxanthin de-epoxidase.
XX
XX
KW Violaxanthin de-epoxidase; VDE; light; photosensitivity;
KW Photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
KW xanthophyll; tobacco.
XX
OS Nicotiana tabacum cv. xanthi.
XX
XX
FH Key Location/Qualifiers
FT Peptide 1..134
FT /label= Transit_peptide
FT Protein 135..478
FT /label= Mat_protein
FT Peptide 135..147
FT /note= "Claim 8"
FT Domain 135..206
FT /label= Cys-rich_domain
FT Domain 385..478
FT /label= Highly-charged_domain
FT Misc-difference 141
FT /note= "conserved Cys residue"
FT Misc-difference 143
FT /note= "conserved Cys residue"
FT Misc-difference 148
FT /note= "conserved Cys residue"
FT Misc-difference 155

FT Misc-difference /note= "conserved Cys residue"
 FT 161 /note= "conserved Cys residue"
 FT Misc-difference 167 /note= "conserved Cys residue"
 FT 171 /note= "conserved Cys residue"
 FT Misc-difference 180 /note= "conserved Cys residue"
 FT 184 /note= "conserved Cys residue"
 FT Misc-difference 190 /note= "conserved Cys residue"
 FT 206 /note= "conserved Cys residue"
 FT Misc-difference 252 /note= "conserved Cys residue"
 FT 382 /note= "conserved Cys residue"
 FT W09717447-A2.
 PD 15-MAY-1997.
 XX 07-NOV-1996; 96WO-US18291.
 XX 06-AUG-1996; 96US-0023502.
 PR 07-NOV-1995; 95US-0006315.
 XX (CALJ) CALGENE INC.
 XX Bugos RC, Rockholm DC, Yamamoto HY;
 PI WPT; 1997-281036/25.
 DR N-PSDB; AAT66242.
 XX DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light
 XX Disclosure; Fig 2; 4lpp; English.
 XX The 55 kba violaxanthin de-epoxidase (VDE) (AAW09875) of tobacco
 CC catalyses the de-epoxidation of violaxanthin to zeaxanthin and
 CC antheraxanthin. This system, termed energy dependent non-radiative
 CC energy dissipation or non-photochemical fluorescence quenching,
 CC reduces the quantum efficiency of photosystem II (PSII), helping to
 CC prevent PSII over-reduction and photoinhibitory damage. The amino
 CC acid sequence of the VDE was deduced from an isolated cDNA clone
 CC (AAT66242). VDE nucleic acids (see also AAT66241, AAT66243), in sense
 CC or antisense orientation, can be used in genetic constructs to
 CC modify VDE levels in plants. Increased levels result in the plant
 CC being tolerant of increased light and therefore more productive
 CC and/or more resistant to disease. Underexpression of VDE increases
 CC photosynthetic efficiency under low light. The photosensitivity of
 CC a range of crops, trees and ornamentals can be modified.
 XX SQ Sequence 478 AA;

Query Match 34.1%; Score 2554; DB 18; Length 478;
 Best Local Similarity 99.8%; Pred. No. 3.9e-191;
 Matches 476; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 474 MALAPHNFIANHETIKYVGSGLPGHKFRSGWEDYFGSIVVAKICSSRRIPRYFRKSP 533
 Db 1 malaphnflanhetikyyvgsklpgkfrfsgwedfygsvivvakicssrripryfrksp 60
 QY 534 RICCGDSRLQLFSGHKHNLSPAHSINQVPGKNSGCKFPKDVLMVWEKQGQAKTAI 593
 Db 61 riccgldsrvgqlfshgkhnlsphasinqvpkgnsqgckfpkdvalmvwekvgqfaktai 120
 QY 594 VAIFILSVASKADAVDAKTCCLLEKCRLEAKLCISNPACAAVACLOTQNNRDETEC 653
 Db 1 vaifilsvaskadavdaaktccllekrleaklcisnpacaaavacloctqnnrdepec 171

Db 121 vaifilsvaskadavdaaktccllekrleaklcisnpacaaavacloctqnnrdepec 180
 QY 654 QIKCGDLFENSVDENECASVRKKCVPRKSDVGFPPDPDSVLVQKDFMDFSGKWEIT 713
 Db 181 qikcgdlfensvdfenecavsrkkcvprksdvdfppdpdpsvlvqkdfmdfsgkweif 240
 QY 714 RGLNPTFADFQCLHEFHTEENKLVGNLSWRIRTPDGGFFTRSAVQKFWODPKYPGILYN 773
 Db 241 rglntpfadfcqlhefhteenkvlgnlswrirtpdggfftrsavqkfvqdpkypgilyn 300
 QY 774 HDNEYLLYQDDWYILSSKVENSPEDYIFVYKGRNDAMDGYGGSVLYTRSAVLPEIPE 833
 Db 301 hdneylllyqddwyilsskvenspedyifvykgndawdgyggsvlytrsavlpeiipe 360
 QY 834 LOTAAQKVGGRDENTFIKTNTCGPEPLVERLEKKVSGERTIIEVEIEEVEEKVDRK 893
 Db 361 lqtaaqkvgrdntfiktntcgppeplverlekkvsgertiikeveieeveekvdrk 420
 QY 894 EYTLFSKLFEGFKELQRDEENFLRELSKEEMDVLDGLKMEATEVEKLFGRALPIRKL 950
 Db 421 evtlfsklfegfkqlqrdeenflreelskeemdvlldglkmeateveklfgralpirkl 477

RESULT 2
 AAW09874
 ID AAW09874 standard; Protein: 473 AA.
 AC AAW09874;
 XX 28-JUL-1997 (first entry)
 DE Romaine lettuce violaxanthin de-epoxidase.
 XX Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 KW xanthophyll; lettuce.
 XX Lactuca sativa L. cv. romaine.
 OS Key Location/Qualifiers
 FH Peptide 1..125
 FT /label= Transit_peptide
 FT Protein 126..473
 FT /label= Mat_protein
 FT Peptide 126..138
 FT /note= "Claim 8"
 FT Domain 126..197
 FT /label= Cys-rich_domain
 FT Peptide 218..231
 FT /label= Lipocalin_signature
 FT Domain 376..473
 FT /label= Highly-charged_domain
 FT Peptide 265..272
 FT /label= Tryptic_peptide-11
 FT Peptide 275..289
 FT /label= Tryptic_peptide-21
 FT Peptide 341..353
 FT /label= Tryptic_peptide-15
 FT Misc-difference 132 /note= "conserved Cys residue"
 FT Misc-difference 134 /note= "conserved Cys residue"
 FT Misc-difference 139 /note= "conserved Cys residue"
 FT Misc-difference 146 /note= "conserved Cys residue"
 FT Misc-difference 152 /note= "conserved Cys residue"
 FT Misc-difference 158 /note= "conserved Cys residue"
 FT Misc-difference 162 /note= "conserved Cys residue"
 FT Misc-difference 162 /note= "conserved Cys residue"
 FT Misc-difference 171

FT Misc-difference 175 /note= "conserved Cys residue"
 FT FT Misc-difference 190 /note= "conserved Cys residue"
 FT FT Misc-difference 197 /note= "conserved Cys residue"
 FT FT Misc-difference 243 /note= "conserved Cys residue"
 FT FT Misc-difference 373 /note= "conserved Cys residue"
 FT FT Misc-difference 473 AA;
 XX WO9717447-A2.
 XX 15-MAY-1997.
 XX 07-NOV-1996; 96WO-US18291.
 XX 06-AUG-1996; 96US-0023502.
 XX 07-NOV-1995; 95US-0006315.
 XX (CALJ) CALGENE INC.
 XX Bugos RC, Rockholm DC, Yamamoto HY;
 XX WPI; 1997-281036/25.
 XX N-PSDB; AAT66241.
 XX DNA encoding plant violaxanthin de-epoxidase - used to modify the
 XX sensitivity of a plant to light
 XX Example 1; Fig 1; 4lpp; English.
 XX The 55 kba violaxanthin de-epoxidase (VDE) (AAW09874) of romaine
 CC lettuce catalyses the de-epoxidation of violaxanthin to zeaxanthin
 CC and antheraxanthin. This system, termed energy dependent
 CC non-radiative energy dissipation or non-photochemical fluorescence
 CC quenching, reduces the quantum efficiency of photosystem II (PSII)
 CC helping to prevent PSII over-reduction and photoinhibitory damage.
 CC The amino acid sequence of the VDE was deduced from an isolated
 CC cDNA clone (AAT66241). VDE nucleic acids (see also AAT66242-43), in
 CC sense or antisense orientation, can be used in genetic constructs
 CC to modify VDE levels in plants. Increased levels result in the
 CC plant being tolerant of increased light and therefore more
 CC productive and/or more resistant to disease. Underexpression of
 CC VDE increases photosynthetic efficiency under low light. The
 CC photosensitivity of a range of crops, trees and ornamentals can be
 CC modified.
 XX SQ Sequence 473 AA;
 Query Match 33.3%; Score 2498; DB 18; Length 473;
 Best Local Similarity 99.6%; Pred. No. 9.2e-187;
 Matches 471; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MALSLHTVFLCKEALNLYARSPCNERHRSQOPPTNIIMMKIRSNNGYFNSFLFTSYK 60
 DB 1 MALSLHTVFLCKEALNLYARSPCNERHRSQOPPTNIIMMKIRSNNGYFNSFLFTSYK 60
 QY 61 TSFSDSSSHCKDKSQICSDITSFEEIQRLDKRGMTLILEKQWQFIQLAIVLVCTFVIV 120
 DB 61 TSFSDSSSHCKDKSQICSDITSFEEIQRLDKRGMTLILEKQWQFIQLAIVLVCTFVIV 120
 QY 121 PRYDAVDALKTCACLLKECRTELAKCIANPSCAANVACLTQCNRPDTEQIKGDLFE 180
 DB 121 PRYDAVDALKTCACLLKECRTELAKCIANPSCAANVACLTQCNRPDTEQIKGDLFE 180
 QY 181 NSVVDQFNCAVSRKCKVPRKSDVGEFPVDRNAVQNFNMKDFSGKWYITSGLNPTFDA 240
 DB 181 NSVVDQFNCAVSRKCKVPRKSDVGEFPVDRNAVQNFNMKDFSGKWYITSGLNPTFDA 240
 QY 241 FDCQLHEFHENDKLVGNLTWRIKTLGSGFTTRSAVQTFVQDPDPLPGALYNHNEFLHYQ 300

DB 241 FDCQLHEFHENDKLVGNLTWRIKTLGSGFTTRSAVQTFVQDPDPLPGALYNHNEFLHYQ 300
 QY 301 DDWYILSSQIENKPDYIFVYVYGRNDADWDGYSVYVTRSTPLPESIIPINLQKAASVG 360
 DB 301 DDWYILSSQIENKPDYIFVYVYGRNDADWDGYSVYVTRSTPLPESIIPINLQKAASVG 360
 QY 361 RDNFNFIITDSCGPEPPPLVERLEKTAEGEKLLIKEAVEIEEVEKEKVRDTEMTLF 420
 DB 361 RDNFNFIITDSCGPEPPPLVERLEKTAEGEKLLIKEAVEIEEVEKEKVRDTEMTLF 420
 QY 421 ORLLGFKELQODEENFVRELKKEKEILNELQMEATEVEKLFGRALPIRKLR 473
 DB 421 QRLEGFKEIQDEENFVRELKKEKEILNELQMEATEVEKLFGRALPIRKLR 473
 RESULT 3
 AAW09876
 ID AAW09876 standard; Protein; 462 AA.
 XX AC AAW09876;
 XX 28-JUL-1997 (first entry)
 XX Arabidopsis violaxanthin de-epoxidase.
 XX Violaxanthin de-epoxidase; VDE; light; photosensitivity;
 KW photoprotection; transgenic plant; zeaxanthin; antheraxanthin;
 KW xanthophyll.
 XX Arabidopsis thaliana var. columbia.
 XX Key Location/Qualifiers
 FT Peptide 1..113
 FT Protein /label= Transit_peptide
 FT Peptide 114..462
 FT Peptide /label= Mat_protein
 FT Domain 114..126
 FT Domain /note= "Claim 8"
 FT Domain 114..185
 FT Domain /label= Cys-rich_domain
 FT Domain 364..462
 FT Misc-difference 120 /label= Highly-charged_domain
 FT Misc-difference 122 /note= "conserved Cys residue"
 FT Misc-difference 127 /note= "conserved Cys residue"
 FT Misc-difference 134 /note= "conserved Cys residue"
 FT Misc-difference 140 /note= "conserved Cys residue"
 FT Misc-difference 146 /note= "conserved Cys residue"
 FT Misc-difference 150 /note= "conserved Cys residue"
 FT Misc-difference 159 /note= "conserved Cys residue"
 FT Misc-difference 163 /note= "conserved Cys residue"
 FT Misc-difference 178 /note= "conserved Cys residue"
 FT Misc-difference 185 /note= "conserved Cys residue"
 FT Misc-difference 231 /note= "conserved Cys residue"
 FT Misc-difference 362 /note= "conserved Cys residue"
 XX WO9717447-A2.
 XX 15-MAY-1997.

XX 07-NOV-1996; 96WO-US18291.
 XX 06-AUG-1996; 96US-0023502.
 PR 07-NOV-1995; 95US-0006315.
 XX (CALJ) CALGENE INC.
 PA Bugos RC, Rockholm DC, Yamamoto HY;
 XX WPI; 1997-281036/25.
 DR N-FSDB; AAT66243.
 XX DNA encoding plant violaxanthin de-epoxidase - used to modify the
 PT sensitivity of a plant to light
 XX Disclosure; Fig 3; 4lpp; English.
 XX The violaxanthin de-epoxidase (VDE) (AA09876) of Arabidopsis
 CC catalyses the de-epoxidation of violaxanthin to zeaxanthin and
 CC antheraxanthin. This system, termed energy dependent non-radiative
 CC energy dissipation or non-photochemical fluorescence quenching,
 CC reduces the quantum efficiency of photosystem II (PSII), helping to
 CC prevent PSII over-reduction and photoinhibitory damage. The amino
 CC acid sequence of the VDE was deduced from an isolated cDNA clone
 CC (AAT66243). VDE nucleic acids (see also AAT66241-42), in sense or
 CC antisense orientation, can be used in genetic constructs to modify
 CC VDE levels in plants. Increased levels result in the plant being
 CC tolerant of increased light and therefore more productive and/or
 CC more resistant to disease. Underexpression of VDE increases
 CC photosynthetic efficiency under low light. The photosensitivity of
 CC a range of crops, trees and ornamentals can be modified.
 XX Sequence 462 AA;
 SQ

Query Match 32.2%; Score 2417; DB 18; Length 462;
 Best Local Similarity 99.8%; Pred. No. 1.9e-180;
 Matches 461; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 951 MAVATHGCTSPCHDRIRFFSDGIGRLGTRKTRKINGTFLKLIPPTQSDALRTTGRRSS 1010
 Db 1 mavathgctspchdrirffsdgigrlgtrkringtflklipptqsdalrttgrss 60
 QY 1011 RPLSAFRSGSKGIFDVLVPLSPKNEUKELPAPLLKLVLGVLACAFILVPSADAVDAKTC 1070
 Db 61 rplsafrrsgskgifdvlvplspkneukelkaptllllklvglvacafilvpsadavdalktc 120
 QY 1071 ACLLKGRICELAKCIANPACAANVACIQTCCNRRPDETECOIKGDLFENSVDDEFNECAY 1130
 Db 121 acllkgrielakcianpacaanvaciqtcnrrpdeceqikcgdlfensvvddefnecav 180
 QY 1131 SRKCVPRKSDLGEPFAPDPSVLVQNFNISDFNGKWIYITSGLNPTFDQCLHEPHTGEG 1190
 Db 181 srkcvprksdglgepfapdpssvlvqnfnsdfngkwyitsglnptfdqclhefhteg 240
 QY 1191 DNKLGVNISRKIKTLDGFFTRSAVQKVFQDPNQPGLVYNDNEYLHYQDDWYILSSKIE 1250
 Db 241 dnklgvnlsrriktdlgfftrsavqkvfqpdpnqpqglvynhndneylhyqddwyilsskie 300
 QY 1251 NKPEDIYFYVYGRNDAMWDGYGAVVYTRSSVLPNSIIPLEKAAKSIGRDFSTFTRTDN 1310
 Db 301 nkpediyfyvyrndacdgygavvytrssvlpnsiiplekaaksigrdfstfirtdn 360
 QY 1311 TCGPEPALVRIEKTVEEGRITVKEVEIEEVEVEKEVKGTEMTLFORLAEGNELK 1370
 Db 361 tcgpepalveriektveegeritvkeveieeveevekevekgvtemtllfqlaegfnelk 420
 QY 1371 QDENFVRELKSEMEFFLDRIKMEASEVEKLFCKALPIRKVR 1412
 Db 421 qdeenfvrelskeemefldrikmeaseveklfckalpirkvr 462

RESULT 4
 AAG49996
 ID AAG49996 standard; Protein; 525 AA.
 XX
 AC AAG49996;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 63310.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 09-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.

PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151086.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	01-SEP-1999;	99US-0151930.
PR	23-JUN-1999;	99US-0140354.	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	99US-0140695.	PR	10-SEP-1999;	99US-0153070.
PR	28-JUN-1999;	99US-0140823.	PR	13-SEP-1999;	99US-0153758.
PR	29-JUN-1999;	99US-0140991.	PR	15-SEP-1999;	99US-0154018.
PR	30-JUN-1999;	99US-0141287.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0141842.	PR	20-SEP-1999;	99US-0154779.
PR	01-JUL-1999;	99US-0142154.	PR	22-SEP-1999;	99US-0155139.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144633.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160767.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160815.
PR	22-JUL-1999;	99US-0145089.	PR	22-OCT-1999;	99US-0160815.
PR	22-JUL-1999;	99US-0145192.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160989.
PR	23-JUL-1999;	99US-0145224.	PR	25-OCT-1999;	99US-0161404.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161359.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-01

XX Arabidopsis thaliana protein fragment SEQ ID NO: 3067.
DE Protein identification; signal transduction pathway; metabolic pathway;
XX Hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
KW
OS
XX Arabidopsis thaliana.
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 23-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 03-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 14-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 13-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.

```
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.8%; Score 209; DB 21; Length 522;
Best Local Similarity 21.2%; Pred. No. 8.4e-08;
Matches 122; Conservative 76; Mismatches 191; Indels 186; Gaps 29;

Qy 375 PEPLVERLEKTAEGEGLIKKEAVELEEEVEKEVETEMTLFQRLLEGFKELQODE 434
Db 68 peppv--klialvdkgevsplk--stseevmlhtarrlkwd-----egymivfdd 116
Qy 435 ENFVRELKSEKETLNELQWEATEVEKLGKALPIKRLMALAPHSNFLANHETIKYVG 494
Db 117 e-----ilssndqtalt-lkqelndqtdilvva-----vnnsevs----- 150
Qy 495 SKLPGHKRFSMGWEDYFGSIVVAKRSRRIPRYFRKSPRICGLDSRGLQLFSGHKHNL 554
Db 151 -----nwlgtnsqnkvnmic-----fesspnlmrlgg----- 178
Qy 555 SPAHSINONPKNGSGCKFKPDVLMV-----WEKNG-----QFAKTAIVAIFILSV--- 601
Db 179 tdvsgvskd-----kevtevkvtvedawernnsddirfclliinayirpvpvl 227
Qy 602 ---ASKADAVDALKTCTLLKECRLEAKCISNPACAAANVACLOTQNNRPDETECQKCG 658
Db 228 qnlrsk-----gfstlscmwkncpqiinclldpncrkaiqlnqcs--pvdqvcysrci 280
```

```
Qy 659 DLFENSVDDEFNECAVSRKKCVPRKSDVGDPPDPDPVSVLVQKFDKDFSGK----- 709
Db 281 asyespyfeafslcvlgkhncldeidakipvkpypp-----mtsfrgkelchdtaed 332
Qy 710 -----WFTIRGLNPTFDADFQCLHEFHTEENK-----LVGNLSWR 744
Db 333 lfvglgelewsrvvvaqnpaydqfpcqqlfyrgkqkssfwyepfvqvtllegklvwr 392
Qy 745 IRTPDGGFFTRSAVQKVFQDPKYP-----GILYNH--DNEYLLYQDDWYILSSKVENSP 797
Db 393 rr-----rysvkr-----gkipatfrfsvldngvvsnef-----wtlvd--vsddls 432
Qy 798 DYIFVYKGRNDAMDWDGYSVLYTRSAVLP-ESIPELQTAQKVG-RDFWTEIKTNTC 855
Db 433 wglfhyngaarvagqsytagvlvtpdgsypaekerklsalekcgikewelf-avdncls 491
Qy 856 GPEPLVERLEKKVEGER-----TIKEVEEIEE 885
Db 492 cenppl-----gipqgsrlhsrisileepdseek 520

RESULT 7
AAG06330
ID AAG06330 standard; Protein; 526 AA.
XX AC AAG06330;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 3066.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
```

PR 20-MAY-1999; 99US-01351124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139433.
PR 17-JUN-1999; 99US-0139432.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140355.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.


```

SQ Sequence 1055 AA;
Query Match 2.0%; Score 151; DB 21; Length 1055;
Best Local Similarity 19.3%; Pred. No. 0.0084;
Matches 182; Conservative 119; Mismatches 306; Indels 334; Gaps 43;

QY 128 ALKTCACLLKECRIE-----LAKTANPSAANVACLOTCNNRPDETQCIKCGDLFEN 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57 ailtalafgragrtqraatlkdftkgcsyavqvkmksngedafkselygvgviiie 116
QY 182 SVVDFNECAVSR----KKCVPRKSDVGPFPDRAVAVQNFNM-----KDFSGK 227
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 117 rritesatavldykgkvsrldelre-----lvhefnidvenpcvmsqkags 168
QY 228 WYITSLNPTFDADFQOLHEFHENDKVLGNLTWRITLDGGFFTRSAVOTFVQDPDLP 287
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 syi-----leak-----gn-----sssfllrnlq---qvnldllq 194
QY 288 ALYNHNDNEFLHYQDDWYLLSSQIEN--KPDYIFVYVYGRNDWDGYSVYVTRSP 345
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 195 siyehlktataidv-----elentikpiekeiselrgk-----ikmeqv 234
QY 346 ESITPNLQKAAK-----SVGRDFNFNTTNSCGPEPLVERLEKTAEEGKLLIKE 397
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235 eelaqrlqllkklawswydvgrq-----lqetekiivklkeriptcq 278
QY 398 AVEIEEVEKEVEKVRDT-----EMTLFORLEGFKE-----LQOD 433
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 a-kidweigk-veslrdltlckkaqvacldestamkreiesfhgsaktavrekialqee 336
QY 434 -----ENFVRELSKEKEILNELOMEATEVEKLFGRALP 468
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 337 fnhkcnyvqkikdrvrrlerqvgdineqtmkntgaeseieeklylerevek----- 389
QY 469 IRKLRLMALAPHSN-----FLANHETIKYVYVYVYVYVYVYVYVYVYVYVYVY 510
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 vetlrsrkeeneeflekafegrrkkmehiedmknhgkrqrfitsnindlkkhqtakvta 449
QY 511 FGSIVWAKICSS--RRIPRYFRKSPRICGLDSRGLQFSGHKNLSPAHSINQVPRKNS 569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 450 fggdrvinllqaiernhrrfrkbp-----igpigs----- 479
QY 570 CCKFPKDVALMWKQWQFQAKTAIVAILSVASKADAVDALKTCICLLKECRLELAKCI 629
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 -----hvtlvngnkassveqalgt--llnafivtdhkdsI----- 513
QY 630 SNPACAAVACLOTCNNRPDETQCIKCGDLFENSVDNEFNECAVS--RKCKVPRKSDVGD 688
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 -----tlrgcaneanyrniki-----liydsrprlniprhmvptqehpti 554
QY 689 FVPDPSPVLVQKDFMDKDFSGKWFITRGLNPTF-----DAFDCQLHEFTEENKLV--- 738
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 555 fsvidsd-----nptflnvlvdqsgverqvlaenyegkavafg 593
QY 739 ---GNLSWRIRTPPG--GFFTRSAVOKFVQD--PKYPGILYVNDHEYLLYQDDWILSKVE 793
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 594 krlsnlk-evytlgdymffrgvqvtlpllsrrpsrlcasfdqj---kdleleaskeq 649
QY 794 NSPEDYIFVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 853
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 ne-----lnqcmrrreeneleelkvrlqkhrsgaekvlttkelemhdlk-----n 699
QY 854 TCGPEPPLVERL-BKKVEEGERTIIKEVEEIEEB---VEKVRD--KEVTL-----FSKLFE 903
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 700 tvaae---ieslpssvnelqreimkdldeideakeaflekqlncleakankltalfe 756
QY 904 GFKELORDE-----ENFLRELSEKEMDVLGLKMEATEVEKL 940
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 muresakgeidafeaeenelkkiekd-----lqsaeeaki 791

RESULT 11
```


also necessary for the cell to advance into the next stages of mitosis. The mitosis protein, can be used to control the growth of cells. An anti-mitosis antibody, a mutant or a non-functional analogue of mitosis can inhibit the mitotic cell cycle by preventing the cells from entering the M phase, and over expression of mitosis or its functional equivalent, would inhibit the cycle by preventing cells from leaving the M phase. Antagonists to this protein can be used to control hyperproliferative cells in, (e.g. thyroid hyperplasia, Grave's disease, psoriasis, benign prostatic hypertrophy, Li-Fraumeni syndrome, breast cancer, sarcomas and other neoplasms, bladder cancer, colon cancer, lung cancer and various leukaemias and lymphomas). Reintroduction or supplementation of lost mitosis function by introduction of the protein or nucleic acid encoding the protein into a cell can restore defective chromosome segregation, which is a marker of progressing malignancy. Malignant proliferation of cells can then be halted. The protein can also be used for the detection and diagnosis of hyperproliferative cells.

Sequence 2482 AA;

Query Match 2.0%; Score 147; DB 19; Length 2482;
Best Local Similarity 18.4%; Pred. No. 0.064;
Matches 240; Conservative 188; Mismatches 443; Indels 434; Gaps 60;

QY 379 LVERLEKTAEEGKLLKEAVEIEEVE-----KEVE-KVRDTEMTLFORLEGFKELOQ 432
DB 315 lsetlslekkemsiainkreltelqngtlikeinasinqemknliq-sesfanyid 373
QY 433 DEENFVRELS---KEKEEII-----449
DB 374 ereksiselsdqyqekllilqrceetgnayedlsqykaaqaeksklecllnectslce 433
QY 450 ---NELQMEATEVEKLFGRALP---IRKRLMALPHNSFLANHETIKYVYGSKLPGH----500
DB 434 nrknele-----qikeafakehgefiltkfaaeernqulmleletvqqlrsemtdnqns 489
QY 501 KRFSWGVEDYFGSIWAKICSSRIPIRYFKSPRI-----CCGLDSRGQLGFESHGK 551
DB 490 ksaagglkqemtikeeqnkmqkevndllqeneqlnmkthhecqnlesepri-----543
QY 552 HNLSPAHSINQNPVKGSGCKFPKDVAMVWEKWGOFAKTAIVAFILSVASKADAVDAL 611
DB 544 -----nsvkeresernq-cnfqpmqndlevkeisldsyanaqlvqleamlrnkkelkqese 596
QY 612 KTCTCL-----LKECRLEL-AKCISNP-----ACAAN 637
DB 597 kekeclqhelqtlrldletsnlqdmgsqelsglkdceidaeekeyisgphelstsgndnah 656
QY 638 VAC-LQTCNNRPDETE--COIKGCDLFENSVDPEFNECAVSRKKCV--PRK--SDVGDF- 689
DB 657 lqcsldtmnklnlelekeiceilqaeye--lvtelnd---srsecitatrkaeevgkil 711
QY 690 ----PVPDPSVLVOKFDMKDFSKWIFITRGLNP-----TFDAFDC 725
DB 712 nevkiindsgllhelvedipdgef---geqneqhpvsplapdesnsyehltisdkv 768
QY 726 QLHEFTEENKLVNLSWRITPDGCFTRSAVQKFPVQPKYPGIILYHNDNEVLLYQD--783
DB 769 qmhfaelqek-----flslqsehkllhdqhcq--msskmselqtyvdsi 810
QY 784 --DWYILSSKVENSPEDYIFVYVYKGRND-----AMDGYGGSVLYTRSAVL 826
DB 811 kaenlvstnlrnfqgdvlvkmqglieeglvpslssscvdpdssslsslgdssfy--rall 868
QY 827 PE----SIPELQTAQ----KVGDFNFTFIKTDNCGPEPPL-----VERLEKKVEEGE 873
DB 869 eqtgdmsllslnlegavsanqcsdevfcslqeenlitrketpsapakgveelesicevyr 928
QY 874 RTIIKEVEEIEEVEER---VROKEVTLFSLKLFEG-----FKELQREDEENFLRELSKE 922
DB 929 qsl-----ekleekmesqimknkeiqeleqlsiserqeldclrkqylseneqwgqkltev 984

QY 923 EMDVLGLKMEATEVEKLF-----GRALPIRKLMVAVATH-----CFTS-- 960
DB 985 tiemesklaaekktedlslelevarlqlgldlssrslgldtedalqgrnescdiske 1044
QY 961 -----PCHDRIRFTSSDDIGRLGITRKRIRINGTFLLLKILPPI-----QSADL---1002
DB 1045 htsettertpkhd-vhqiackdaqqdlnldiekitetgalk---ptgecsgeqspdtuue 1100
QY 1003 ----RTTGGRS-----SRPLSAFRSGFSGIFDIVPLPSK-----1033
DB 1101 ppegdktdsgsseciselsfsgpnalvpmdfingqediqlrvktsnenlrlhvied 1160
QY 1034 -----NELKELTAPLLKLVGLACAFILVPSADAVDALKTACALLKGCRIELAKC 1084
DB 1161 rarkvesllnemkelksklhqlveqlmt-----kieac-----ieleki 1199
QY 1085 IANPACAANVACLQTCNNRPDETECOIKGCDLFENSVDPEFNECAVSRKKCVPRKSDLG-1143
DB 1200 v-----gel-----kk-----ensdlse 1212
QY 1144 --EPPAPDPVLQVQNFISDFNGKWYITSGLNPTFDADFDCOLHEFHTE-----GDNKLVGN 1197
DB 1213 kleyfscdhqellqrvtse-----glns-----dlemhadksredigdnvavkn 1258
QY 1198 ISWRIKTLDSGFFTRSAVQKFPVQDPNPQGVLYNHNDNEYLH-----YQDDWYILSKSIENK 1252
DB 1259 dswkerfld-----venelsirsekasihealyleadlevvqtkeklcledknenk 1310
QY 1253 P-----EDYIFVYVYGRNDADWDGCGAVVYTRSSVLPNSIIPELEKRAKSGIRDFSTFI 1306
DB 1311 qkvivcleelsvvtsernql---rgeldtmskktaldqlsekmektqleshqsecl 1367
QY 1307 RTDNTCGPEPALVERIEKTEVEEGERIIVKEVEEIEE---VEKEVEKVGRTMTLFORLA 1363
DB 1368 hciqvaaevktektellqtssdvselkldkthlqeklqslsleqsalsltkcelenqia 1427
QY 1364 EGFNELKODENFVRE-----LSKEEMEFLEDEIK-WEASEVEK 1400
DB 1428 ----qlnkekellvkeselqarisesdyeklnvskaleaalvek 1468

RESULT 14

AAB18324
ID AAB18324 standard; Protein; 1558 AA.

AC AAB18324;

DT 07-NOV-2000 (first entry)

XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:182.

XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine; antimalarial; malaria; protozoacide; infection; insecticide.

XX Plasmodium falciparum.

XX WO200025728-A2.

XX PD 11-MAY-2000.

XX PF 05-NOV-1999; 99WO-US26796.

XX PR 05-NOV-1998; 98US-0107131.

XX PA (HOFF/) HOFFMAN S.

XX PA (CARU/) CARUCCI D.

XX PA (GARD/) GARDNER M.

XX PA (VENT/) VENTER J C.

XX PI Hoffman S, Carucci D, Gardner M, Venter JC;

XX WPI; 2000-365347/31.

XX

QY 617 LLKECRLELAKCINPACAAANVACLOTCNNRPDTECO---IKGDLFENSVDDEFNECA 673
Db 1110 -LKEMFFNLEDFVKS---ESDVTVEELKDEPVOKEVETVSIIEEMENIVDLEE-- 1163
QY 674 VSRKKCVPRKSDVGD--FPVPDPVSVLQKMDKDFSGKWFITRCLNTPFAFCQLHEFH 731
Db 1164 -----EREDLTKMDADAVEESIEISS-DSKEET-----ESI 1193
QY 732 TEENKLVGNLSWRITPDGGFFTSVAVQKVFQDPKYPGILYNHNDNEYLLODDWYILSSK 791
Db 1194 KDKEKDVSVVEEVDND---MDSEVKEVLE---LKNMEELMKDAVEINDITSK 1242
QY 792 -VENSPEYIFVYKGRNDADWDGGSVLTGRSAVLPESI--IPELOTAQKVGROFNFTF 848
Db 1243 LIEETQE-----LNEVEADLIKDMKEKLEKA---LSEDSKEI 1278
QY 849 IKTDNTCGPEPLVERLEKKEVEGE--RTIIEVEEET---EEVEKVRD-----K 893
Db 1279 IDAKD-----DLEKVEIEEHDITTLDEVELKDVVEEKIEKVSOLKLEEDILK 1329
QY 894 EV-----TLFSKLPFGFKELQDENFRLSKEEMDVLGLKMEATEVEKLFGRAL----- 945
Db 1330 EVKEIKESEILDYKELTETIDILEEKEKEIEKHFEKFEAEAEIKDLEADILKEVS 1389
QY 946 -----PIRKMAVATHCFTSPCHDRIRFFSSDDGIGRLGITRKRINGTFLLKIL 994
Db 1390 SLEVEEKKLEEVHELKEVEEHIISGAH-----IKG----- 1421
QY 995 PPIQADLRTTGGRRSRLPSAFRSGFSKGIIDIVPLPSKNELKELTAPLILKLGVGLACA 1054
Db 1422 --LEEDDLEEVDDLGKSILOMLKGMELGMD-----KESLEDVTKL----- 1462
QY 1055 FLIVPSADAVDAKLTACILKGGRIELAKIANPACAAANVACLOTCN--NRPDTECOIK 1112
Db 1463 -----GERVESLKDVLSSALGMDDEQMKTRKKAQRPKLEBVLK 1501
QY 1113 CGDLFENSVDDEFNECAVSRKKCVPRKSDLGEPFAPDPVSVLVQNFNISDFENGWYITSL 1172
Db 1502 -----EEVKEEKKKITRKK-----VRFDIKD-----K 1524
QY 1173 NPTFADFQCLHFEHTGDKNLKVNISWRIKTLDSGFFTRSAVQKVFQDPNPGVLYNHD 1232
Db 1525 EPKDEIVEEMKDEIED-----VEEDIEED----- 1551
QY 1233 NEYLHYODDWYILSSKLENKPEYIFVYGRNDANDWDGGAIVYVTRSSVLNPSIIPLE 1292
Db 1552 -----IEEDKVEDIDEDIDEDIGDKDEVID-----LIVQKE 1583
QY 1293 KAAKSIGRDFSTFRTDNTCGPEPALVERIEKTVEEGERIIVKEVEE-----IEEVEE 1345
Db 1584 KRIEKV-----KAKKKKLEKVEGVSGLKKHVDVMKYVQKIDKVD 1626
QY 1346 KEVEKV--GRTEMTLQRLAEGNELKQDEE-----NFRVRELSKEEMEFLEIKMEASE 1397
Db 1627 KEVSKALESNDVT-----NVLKQODFSKVKNEVKKYKFAAPFISAVAFAASY 1677
QY 1398 VEXLF 1402
Db 1678 VVGFF 1682

RESULT 2

US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022e1 Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700

; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-254-6

Query Match 2.0%; Score 147; DB 1; Length 2482;

Best Local Similarity 18.4%; Pred. No. 0.0058;

Matches 240; Conservative 188; Mismatches 443; Indels 434; Gaps 60;

QY 379 LVERLEKTAEGEKLIIKEAVEIEEVE-----KEVE-KVRDTEMTLQRLLEGFKELOQ 432
Db 315 LSETLSLEKKEKSSIIISLNKREIEELTQENGLKEINASLNQEKMNLIQK-SESFANYID 373
QY 433 DEENFVRELS---KEEKEIL----- 449
Db 374 EREKSISELSDQYQKEKLIIILQRCETGNAYEDLSQYKAAQEKNSKLECLLNCTSLCE 433
QY 450 ---NELQMEATEVEKLFGRALP--IRKLRLMALPHSNFLANHETIKYIVGSKLPGH--- 500
Db 434 NRKNELE---QLKEAFAKEHGFELTKLAPAEARNQNLMLLELTVQOALRSEMTDNQNS 489
QY 501 KRFSGWEDYFGSIVVAKICSSRRIPRYFRKSPRI-----CCGLDSRGLQFLSHGK 551
Db 490 KSEAGGLQKEIMTLKEBQNKMKQEVNDLLOENFOLMKVMKTKHECQNESEPIR----- 543
QY 552 HNLSPAHSINQVPGNSGCKFPKPDVALMVWEKMGQFAKTAIVAIFILSVASKADAVDAL 611
Db 544 -----NSVKERESERNO-CNFKPQMDLEVKEISLDSYNAQLVQLEAMLRNKLKLOESE 596
QY 612 KTCYCL-----LKECRLEL-AKCISNP-----ACAAN 637
Db 597 KEKECLQHELOQTIRGDLSTSLNLDQMOSQEISGLKDCIDEAEKYISGPHELSTSQNDNAH 656
QY 638 VAC-LQTCNNRPDETE--COLKCGDLFENSVDDEFNECAVSRKKCV--PRK--SDVGDF- 689
Db 657 LOCSLQTTMKNLNELEKICEILOAEKYE--LVTELND---SRSECITATRKMAEEVCKLL 711
QY 690 ----PVPDPVSVLVQKFDKDFSGKWFITRGLNP-----TFDAFDC 725
Db 712 NEVKILNDSDGILLHGLVELDIPGGEF---GEQPNQHPVSLAPLDESNSVEHLTSLSKVE 768
QY 726 QLHEFHTEENKLVNLSWRITPDGGFFTSVAVQKVFQDPKYPGILYNHNDNEYLLODDWY 783
Db 769 QMHEAFELQEK-----FLSLOSEHKLHLDHQHCQ--MSSKMSLEQTYVDSL 810
QY 784 --DWYILSSKVENSPEDYIFVYKGRND-----AWDGYGGSVLYTRSAVL 826

US-08-687-080-51
; Sequence 51, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; APPLICATION NUMBER: US/08/687,080
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/592,126
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 4600-0111.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: TRANS. OF RAD50 CDNA (SEQ. 54), NT.
; INDIVIDUAL ISOLATE: 389 TO 4324
US-08-687-080-51

Query Match 1.8%; Score 135.5; DB 2; Length 1312;
Best Local Similarity 19.7%; Pred. No. 0.02;
Matches 124; Conservative 95; Mismatches 218; Indels 191; Gaps 24;
QY 382 RLKTAEEG-----KLIIKAEVIEEVEKEVVRDTEMTLFRLLGFKELQDDENF 437
Db 398 KLVREGEAGTAKNLMNDFAEKETLKQKQIDEIRD-KKTGLGRIELKSEI----- 449
QY 438 VRLSKEEKILNLOMEATEVEKLGRLPRLKRLMALAPHSNF-----ANHETIKY 491
Db 450 ---LSRKQ-----NELKNVYELQQLGEGSDRIELDLQELIKAEKLSKAEKNSVETLKM 502
QY 492 YVGSKLPLGHKRFSGWEDYFGSVIVAKICSSRRIPRYFRKSPRICGLDSRGLQFLSHCK 551
Db 503 EVIS-LQNEK-----ADDLTLRK-----LDQMEQL-NH--- 530
QY 552 HNLSPAHSINQVPGNSCKPKPKVALMWKVGQFAKTAIVAFILSVASKADAVDAL 611
Db 531 -----HTTTRTQMELMTKDKADKQIKRSRSHDELTSLGYF-----PNKQLEDWL 580
QY 612 KTCCTLLKCEKLELAKICSNPACAAVACLOTNNPPDETCQIK-----CGDLFE 662
Db 581 HSKSKEINQTRDLAKL--NKEIASSEQNKNHINNELKREKQSLSYEDKLFDPVCGSQDF 638
QY 663 NSVYDFNECAVSRKCKVPRKSDVGVDPVPDPVSVLVQKDFMDKDFSGKWFITRGLNPTFDA 722

Db 639 ESDRLKE-----EIEKSKQRAMLAGATAVYSQ 668
QY 723 FDCQLHEFHEENKLVGNLSWRIRTPDGGFFTRSAVQKQVDPKYPGYLYNHNEYLQY 782
Db 669 FITQL-----TDENQSCCPVQQRV-----FQTEAELQEVISD----- 700
QY 783 DDWYILSSKVENSPEDYIFV--YYKGRNDAMDGYGGSVLYTRSAV-LPESITPELQTAQ 839
Db 701 -----LQSKRLAPDKLKSTSELAKKKEKRDEMLGLVPMRQSIIDLKKEKIPELRNKLQ 755
QY 840 KVGDRDNTFIKD-----NTCGPEPLVERLEKKVVEGERTIKF-- 879
Db 756 NVNRDIQR-LKNDIEEQETILLGTIMPEESAKVCLTDTIMERFQMELKDVVERKIAQAA 814
QY 880 -----VEEIEEVEKVRDKEVTLFSLKLFEGFKELQDERNF-----LRELSKEE 923
Db 815 KIQGIDLDRTVQVQVQKQKQKLDTVSSKIELNRLKIQDQEQIQHLKSTTNELAKSEK 874
QY 924 MDVLDGLKME-----ATEVEKLF 941
Db 875 LOISTNLRQROLEEQTVELSTEVSQSLY 902

RESULT 5
US-08-021-601-2
; Sequence 2, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Arora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-021-601-2

Query Match 1.8%; Score 132; DB 1; Length 776;
Best Local Similarity 17.2%; Pred. No. 0.017;
Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

[illegible]

RESULT

```

US-08-082-849B-2
Sequence 2. Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Aroza, Vagene
APPLICANT: Singh, Vandendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend
STREET: Two Embarcadero Center

```

Db 511 AGYLENG-----KLILQNRNIGLEIKDVQ----- 533
QY 1223 NOPGVLYNHNEYLHYQDDWYLSSKIENKPEDYIFVYGRNDWDGCGAVVYTR--S 1280
Db 534 ----LIKSEKEYIRI-DAKVVPSKIDTKIOE-----AQLNINQEWNKALGLPKYTKLIT 584
QY 1281 SVLPNSIIPELEKAASIGRDFSTFIRTDNCGPEPALVERIEKTVERGE--RIIVKEV-- 1337
Db 585 FNVHNRYSNIVESAYLILNWKNNIQSD-----LIKVTNYLVDGNGRFFVTDITL 636
QY 1338 -----BEIEBEVEKEVKGRTMTLQRLAEGFNLKODENFVRE----- 1379
Db 637 PNIAEQYTHQDEIYEQVHSGKLYVPESRILLHGFSKGV-ELRNDSEGFIEFHGAHVDDY 695
QY 1380 ----LSKEEME-----FLDEIKMEASEV 1398
Db 696 AGYLLDKNQSDLVYNSKKFIDIFKEGNSL 725

RESULT 7

PCT-US94-01624-2

; Sequence 2, Application PC/TUS9401624

; GENERAL INFORMATION:

; APPLICANT: Leppla, Stephen H.

; APPLICANT: Klimpel, Kurt R.

; APPLICANT: Arora, Naveen

; APPLICANT: Singh, Yogendra

; APPLICANT: Nichols, Peter J.

; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND

; TITLE OF INVENTION: RELATED METHODS

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW

; STREET: Steuart Street Tower, 20th Floor, One Market

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94105

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/01624

; FILING DATE: June 25, 1993

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Weber, Kenneth A.

; REGISTRATION NUMBER: 31,677

; REFERENCE/DOCKET NUMBER: 15280-115

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 543-9600

; TELEFAX: (415) 543-5043

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 776 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; PCT-US94-01624-2

Query Match 1.8%; Score 132; DB 5; Length 776;
Best Local Similarity 17.2%; Pred. No. 0.017;
Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

QY 733 EENKLVGNLSWRIRTPDGGFFTRSAVQKFGVQDPKPGILYNHD---NEVLLYQDDWY--I 787

Db 72 EMYKAIGG---KIYIVDGDITKHISLEALSEDKKIKDIYKDALHHEHVYAKGEPV 128

QY 788 LSSKVENSPEDYIFVYVYKGRNDWDGCGVSVLPESIPELOTAQOKVGRDENT 847
Db 129 L-----VLOSSEDIYVENTEKALN-----VYIEGKILSRDILSKINOPYOKFLDVLNT 176
QY 848 FIKTDNTCG-----PEPPLIVERLEKKVBERGERTIILK----- 878
Db 177 IKNASDSGDQLLFTNQLKEHPTDFSEVFLEQNSNEVQVFAKAFAYIYEPQHRDVLQLY 236
QY 879 -----EVEETEEREVEKVRUKEVTLFSLKLPFGFKE----- 907
Db 237 APAFNMYDKFNQDEINLSLEELKQDRLSRYEKWEKIKQHYQHWSDSLSSEGRGLLKL 296
QY 908 ---LQDDEENFRELSEKEMDVLGLKMEATEV---EKLGRALPI-----RKL 950
Db 297 QPIEPFKDDIHLISLQSEKEKLLKRIQIDSSDFLSTEEKEFLKQIDIRDSLSEEEKL 356
QY 951 MAVATHCFTSPCHDRIRFFSSDDGIGRIGITKRKRINGTFLKILPPIQSAD----LRTTG 1006
Db 357 L-----NRIOQVDSNP-----LSEK---EKEFLKKLKDIOFYDINQRLQDTG 396
QY 1007 GRSSRPL-----SAFRSGFSKGIQFIDIVPLPSKNELKELTAPLLKLVLG 1049
Db 397 GLIDSPSINLVRKQYKRIQIDALLHQSIGSTLYNKIYLYENNMINNLATL----- 450
QY 1050 VLACAFLIIVPSADAVDAKLTACALLKGCRIELAKCIANPACAANVACLQTCNNRPDETEC 1109
Db 451 -----GADLVDS----- 458
QY 1110 QIKCGDLFENSVDDE--FNECAVSRKKCVPRK---SDIGEPFAPDPSPVLVQVNFNISDFNG 1164
Db 459 -----DNTKINRGIFNEFKKFKYSSISSNYMIVDINERPALDNERLKWRIQLSPDTR 510
QY 1165 KWIITSGLNPTFADFQCLHEPHEGDNKLV--GNISIRIKTLDSDGFTFRSAVQKFGVQDP 1222
Db 511 AGYLENG-----KLILQNRNIGLEIKDVQ----- 533
QY 1223 NOPGVLYNHNEYLHYQDDWYLSSKIENKPEDYIFVYGRNDWDGCGAVVYTR--S 1280
Db 534 ----LIKSEKEYIRI-DAKVVPSKIDTKIOE-----AQLNINQEWNKALGLPKYTKLIT 584
QY 1281 SVLPNSIIPELEKAASIGRDFSTFIRTDNCGPEPALVERIEKTVERGE--RIIVKEV-- 1337
Db 585 FNVHNRYSNIVESAYLILNWKNNIQSD-----LIKVTNYLVDGNGRFFVTDITL 636
QY 1338 -----BEIEBEVEKEVKGRTMTLQRLAEGFNLKODENFVRE----- 1379
Db 637 PNIAEQYTHQDEIYEQVHSGKLYVPESRILLHGFSKGV-ELRNDSEGFIEFHGAHVDDY 695
QY 1380 ----LSKEEME-----FLDEIKMEASEV 1398
Db 696 AGYLLDKNQSDLVYNSKKFIDIFKEGNSL 725

RESULT 8

US-08-404-531B-28

; Sequence 28, Application US/08404531B

; Patent No. 5863724

; GENERAL INFORMATION:

; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela

; APPLICANT: Thomas, Gilbert Cote, and Robert Gagel

; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonyleurea Receptor

; Patent No. 5863724

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &

; ADDRESSEE: NO. 5863724ris

; STREET: One Liberty Place 46th. Floor

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

Db 846 STPMIVLDDPFSALDVHLSHLMQAGILELLRDKRTVVLVTHKLYLPHA-DWIIAMK 904
QY 1248 KIENKPEDYIFVYGRNDAMDGYGGAVVYTRSSVLPNSIPELEKAAKSIGRDFSTFIR 1307
Db 905 DGTIQREGTLKDFORSECOLFEHW-----KTLNRQDOELEK-----ETVME 946
QY 1308 TDNCGPEPALVERIEKTEVEGERIIVKEVEIEEVEEKE 1347
Db 947 RK---APEPS---QGLPRAMSRDGLLLDEDEEEEAASE 981

RESULT 10
US-08-488-546A-28
; Sequence 28, Application US/08488546A
; Patent No. 6054313
; GENERAL INFORMATION:
; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
; APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonyleurea Receptor
; Patent No. 6054313
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 6054313ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,546A
; FILING DATE: 07-JUNE-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/404,531
; FILING DATE: 15-MARCH-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: BYLR-0026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1498 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-488-546A-28

Query Match 1.7%; Score 124; DB 3; Length 1498;
Best Local Similarity 18.8%; Pred. No. 0.24;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;
QY 868 KVEEGERTIIEVEEIEEVEKVRDKVTLFSLKIFGEK--ELQRDEENFLRSLKEEMD 925
Db 480 KLSQAORTLL-----EYSNERLQKT---NEMLRGIKLLKYAWENIFCSRVEKTRRK 528
QY 926 VLGLKMEA--TEVEKLFGRALPIRKLMAVATHCFTSPCHDRIRFFSSD-----GIGRL 978
Db 529 EMTSLRAFAYVTSISIFMTAIPAAVLITFV-----GHVSFFKESDFSPVAFASL 580
QY 979 GITRKRINGTFFLL-----KILPIQ-----SADLRTTGGRRSRPLSAFRSGSKG 1023
Db 581 SLPHILVTLPLFLSSVVRVTRVALSVQKLSEFLSAEIREEQCAPREPAPQQA----G 636

QY 1024 IFDIIVPL-----PSKNELKELTAPLLLLKLVGLACAFLIIVPSADAVDAKTCACLLKG 1076
Db 637 KYQAVPLKVVNKRFAREVRDLGLPQ-----RTPSTDG-DADNFCVQIIIGG 684
QY 1077 C-----RIELAKCIANPACAAANVACLOTCNNRNPDETECIKCGDLF 1117
Db 685 FFTWTPDGIPTLSNITIRIPRGQLTWIVGVCGKSSLLLATIG-----EMQKYSGAVF 738
QY 1118 ENSVYDEFNECAVSRKCVPRKSDLGCEFPADPP-----SVLVQNFNI-SDF 1162
Db 739 WNSLPD-----SEGRPOOPRAGDSGRFCQEQPRCPGASQKQKPLLNATVEENITFESPF 793
QY 1163 NGKWY-----ITSGLNPTFDADFQCLHFEHFTGDKNKLGVNISWRIKT----- 1204
Db 794 NKQRYKMVIEACSLQDDIDL-----PHGDTQIQIGERGINLSTGGQRPDQCRPEPST 845
QY 1205 -----LDSGF-----FTRSAVOKFVQDPNQPGVLYNHNDNEHYLHYQDDWYILSS 1247
Db 846 STPMIVLDDPFSALDVHLSHLMQAGILELLRDKRTVVLVTHKLYLPHA-DWIIAMK 904
QY 1248 KIENKPEDYIFVYGRNDAMDGYGGAVVYTRSSVLPNSIPELEKAAKSIGRDFSTFIR 1307
Db 905 DGTIQREGTLKDFORSECOLFEHW-----KTLNRQDOELEK-----ETVME 946
QY 1308 TDNCGPEPALVERIEKTEVEGERIIVKEVEIEEVEEKE 1347
Db 947 RK---APEPS---QGLPRAMSRDGLLLDEDEEEEAASE 981

RESULT 11
US-08-404-531B-6
; Sequence 6, Application US/08404531B
; Patent No. 5863724
; GENERAL INFORMATION:
; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
; APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonyleurea Receptor
; Patent No. 5863724
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 5863724ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/404,531B
; FILING DATE: 15-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: BYLR-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1581 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-404-531B-6

Query Match 1.7%; Score 124; DB 2; Length 1581;
Best Local Similarity 18.8%; Pred. No. 0.26;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

QY 868 KVEGERTIIKEVEIEEVEKVRDEKVTLSKLFEGFK--ELQRDEENFLRELSKEEMD 925
DB 480 KLSQAQRTTL-----EYSNERLKQT---NEMLRGKLLKLYAWENIFCSRVEKTRRK 528
QY 926 VLDGLKMEA--TEVEKLFGRALPIRKLMVATHCTSPCHDIRFFSSDD-----GIGRL 978
DB 529 EMTSLRAFAVYTSISIFMNTAIPAAVLITFV-----GHVSFFKESDFSVAFAFL 580
QY 979 GTRKRINGTFL-----KILPPIQ-----SADLRTTGGRSRPLSAFRSFGSK 1023
DB 581 SLFHILVTLPLLSVVSRVTKALVSQKLSAEIRFEEQCAPREPAQOGA-----G 636
QY 1024 IFDIVPL-----PSKNELKELTAPLLKLVGLACAFIVPSADAVDAKTCACLLKG 1076
DB 637 KYQAVPLKVVNRKRPAREEVRDLGLPQ-----RLTFSTDG-DADNFCVQIIG 684
QY 1077 C-----RIELAKCIANPACAAVACLOTNNRPDETECOIKCGDLF 1117
DB 685 FFTWPDGIPTLNITIRPGQLTMIVGQVCGKSSLLLATLG-----EMOKVSGAVF 738
QY 1118 ENSVVDENECAYSRKCKVPRKSDLGFEFPADP-----SVLVQNFNI-SDF 1162
DB 739 WNSLPD-----SEGRRPQOPRAGDSGRFCQEQPCGYASQKPWLLNATVEENITFESPF 793
QY 1163 NGKWY-----ITSGLNPTFAFDCQLHEFTGDKNLVGNISWRIKT-----1204
DB 794 NKQRYKMWIEACSLQPDIDL-----PHGDOTQIGERGINLSTGGQRPDQCRPEPST 845
QY 1205 -----LDGSGF-----FTRSAVKFQDPNPGVLYNHDNEYLHYQDDWYILSS 1247
DB 846 STPMIVLDDPFSALDVHLSHLMQAGILELLRDKRTVVLVTHKLYLPHA-DWIIAMK 904
QY 1248 KIENKPEYIFVYGRNDADWDGAGVYVTRSVLPNSIIPLEKAAKSGRDFSTFIR 1307
DB 905 DGTIQRGTLKDFORSECOLFEHW-----KTLNRQDQOLEK-----ETVME 946
QY 1308 TDNTCGPEPALVERIEKTVEEGERIIVKEVEIEEVEKE 1347
DB 947 RK---APEPS---QGLPRAMSSRDGLLDEDEEEEAASE 981

RESULT 12
US-08-476-900A-6
; Sequence 6, Application US/08476900A
; Patent No. 6031150
; GENERAL INFORMATION:
; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson
; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonyleurea Receptor
; Patent No. 6031150
; TITLE OF INVENTION: and Method of Detecting Persistent Hyperinsulinemic Hypoglycemia
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESS: NO. 6031150
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476.900A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: BYLR-0027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1581 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-476-900A-6

Query Match 1.7%; Score 124; DB 3; Length 1581;
Best Local Similarity 18.8%; Pred. No. 0.26;
Matches 109; Conservative 72; Mismatches 221; Indels 178; Gaps 26;

QY 868 KVEGERTIIKEVEIEEVEKVRDEKVTLSKLFEGFK--ELQRDEENFLRELSKEEMD 925
DB 480 KLSQAQRTTL-----EYSNERLKQT---NEMLRGKLLKLYAWENIFCSRVEKTRRK 528
QY 926 VLDGLKMEA--TEVEKLFGRALPIRKLMVATHCTSPCHDIRFFSSDD-----GIGRL 978
DB 529 EMTSLRAFAVYTSISIFMNTAIPAAVLITFV-----GHVSFFKESDFSVAFAFL 580
QY 979 GTRKRINGTFL-----KILPPIQ-----SADLRTTGGRSRPLSAFRSFGSK 1023
DB 581 SLFHILVTLPLLSVVSRVTKALVSQKLSAEIRFEEQCAPREPAQOGA-----G 636
QY 1024 IFDIVPL-----PSKNELKELTAPLLKLVGLACAFIVPSADAVDAKTCACLLKG 1076
DB 637 KYQAVPLKVVNRKRPAREEVRDLGLPQ-----RLTFSTDG-DADNFCVQIIG 684
QY 1077 C-----RIELAKCIANPACAAVACLOTNNRPDETECOIKCGDLF 1117
DB 685 FFTWPDGIPTLNITIRPGQLTMIVGQVCGKSSLLLATLG-----EMOKVSGAVF 738
QY 1118 ENSVVDENECAYSRKCKVPRKSDLGFEFPADP-----SVLVQNFNI-SDF 1162
DB 739 WNSLPD-----SEGRRPQOPRAGDSGRFCQEQPCGYASQKPWLLNATVEENITFESPF 793
QY 1163 NGKWY-----ITSGLNPTFAFDCQLHEFTGDKNLVGNISWRIKT-----1204
DB 794 NKQRYKMWIEACSLQPDIDL-----PHGDOTQIGERGINLSTGGQRPDQCRPEPST 845
QY 1205 -----LDGSGF-----FTRSAVKFQDPNPGVLYNHDNEYLHYQDDWYILSS 1247
DB 846 STPMIVLDDPFSALDVHLSHLMQAGILELLRDKRTVVLVTHKLYLPHA-DWIIAMK 904
QY 1248 KIENKPEYIFVYGRNDADWDGAGVYVTRSVLPNSIIPLEKAAKSGRDFSTFIR 1307
DB 905 DGTIQRGTLKDFORSECOLFEHW-----KTLNRQDQOLEK-----ETVME 946
QY 1308 TDNTCGPEPALVERIEKTVEEGERIIVKEVEIEEVEKE 1347
DB 947 RK---APEPS---QGLPRAMSSRDGLLDEDEEEEAASE 981

RESULT 13
US-08-488-546A-6
; Sequence 6, Application US/08488546A
; Patent No. 6054313
; GENERAL INFORMATION:
; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
; APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonyleurea Receptor
; Patent No. 6054313
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &

```

/ ADDRESS:  No. 6054313ris
/ STREET:  One Liberty Place 46th. Floor
/ CITY:    Philadelphia
/ STATE:   PA
/ COUNTRY: USA
/ ZIP:     19103
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Vers
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/488,546A
/ FILING DATE: 07-JUNE-1995
/ CLASSIFICATION: 800
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/404,531
/ FILING DATE: 15-MARCH-1995
/ CLASSIFICATION: 800
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Beardell, Lori Y.
/ REGISTRATION NUMBER: 34,293
/ REFERENCE/DOCKET NUMBER: BYLR-0026
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 215-568-3100
/ TELEFAX: 215-568-3439
/
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1581 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
/ US-08-488-546A-6

```

[illegible]

```

Db      905  DGTQREGTLKDFRSECOLFEHW-----KTLNMRQDQLEK-----ETVME 946
QY      1308  TDNTCPALVARIKTVGEGRIIVKEVEIEEVEKE 1347
           ||| : : : : : ||| ||| |||
Db      947  RK--APEPS--QGLPRAMSSRDGLLLDDEEEEAASE 981

RESULT 14
US-08-574-959A-9
; Sequence 9, Application US/08574959A
; Patent No. 5962224
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Young, Ratna K. Vadlamudi
; APPLICANT: Jaekyoon Shin, Insil Young, Ratna K. Vadlamudi
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/574,959A
; FILING DATE: 19-DEC-95
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: DFN-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 905 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-574-959A-9

```

```

Query Match          1.6%; Score 119; DB 2; Length 905;
Best Local Similarity 31.0%; Pred. No. 0.28;
Matches 39; Conservative 15; Mismatches 34; Indels 38; Gaps 4;

QY   1284 PMSIPELEKAAKSIGRDFSTFTDNTCGEPALVERI-----BKTVEEGERIV 1334
      |::|||               -|::|||::|::|::|::|::|::|::|::|::|::|
Db    623 PPQLYPE-----GTPGGGP-PALEDLTVINSGSDFEFEEGESEEE 665

QY   1335 KEVEIEEIVEKEVKVGRTEMTLFQRAGFVELKDEENFVRELSKEMEFLEIKME 1394
      ||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    666 EEEEEEEEEEEEE-----EDFEEEEDEEEFYEEEEEEFEFFEEE 713

QY   1395 ASEVEK 1400
      |::|
Db    714 EGEELE 719

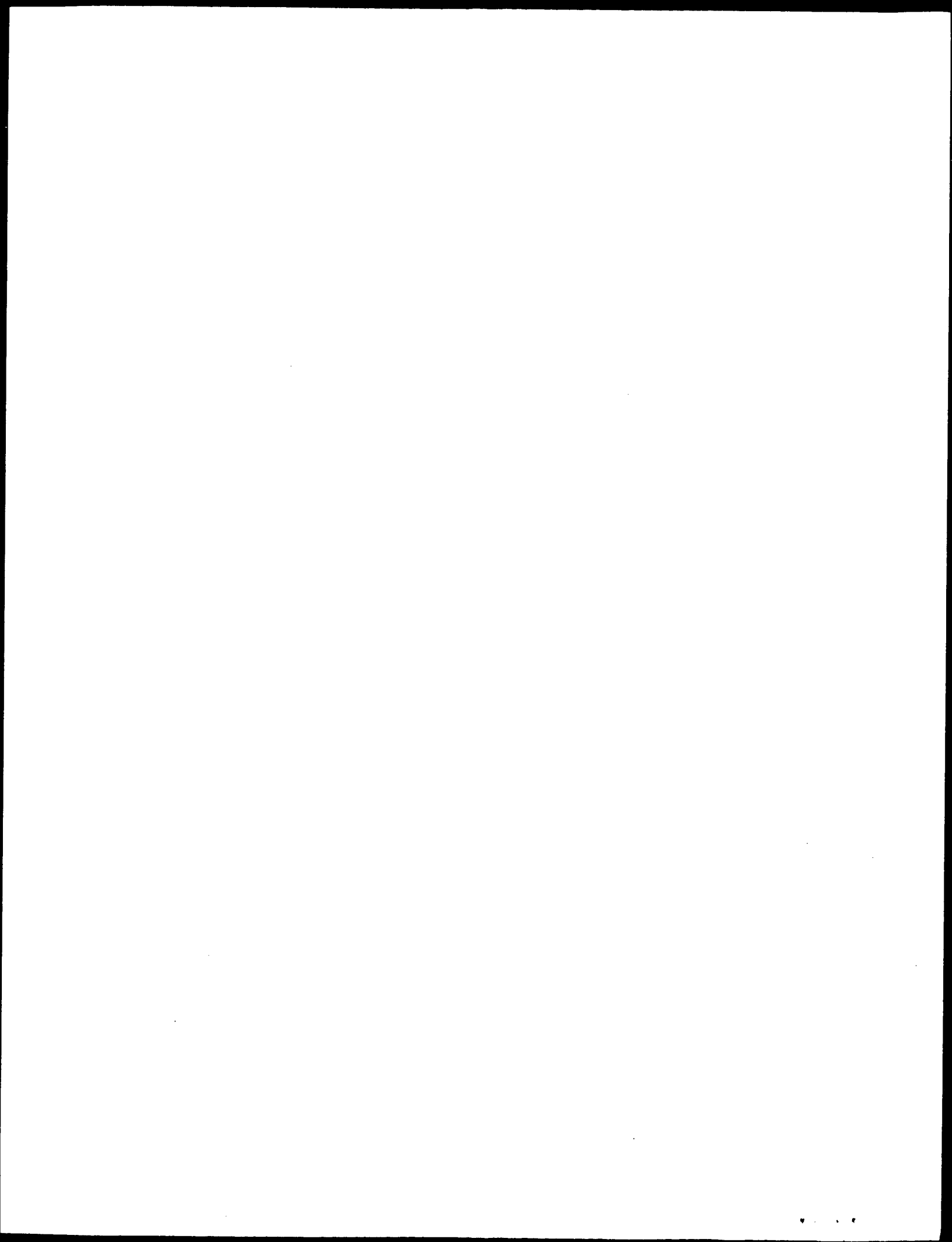
RESULT 15
US-09-357-014-9
; Sequence 9, Application US/09357014
; Patent No. 6291645
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Joong, Ratna K. Vadlamudi
; and Jack L. Strominger
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES

```

AND USES THEREFOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/357,014
FILING DATE: 19-Jul-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/574,959
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: DFN-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 905 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-357-014-9

Query Match 1.6%; Score 119; DB 4; Length 905;
Best Local Similarity 31.0%; Pred. No. 0.28;
Matches 39; Conservative 15; Mismatches 34; Indels 38; Gaps 4;
QY 1284 PNSTIPELEKAAKSGRDFSTFIRDTNCGPEPALVERI-----EKTVEEGERIIV 1334
| :||| | ||||| : |||||
Db 623 PPQLVPE-----GTPGGGP-PALEEDLTVININSSDEEEEGEEEEE 665
QY 1335 KEVEIEEEVEKEVKGRTEMTLQRLAEGFNELKQDEENFVRLSKEMEFLDEIKME 1394
: ||||| :||| : ||||| : ||||| :
Db 666 EEEEEEEEEEEEEE-----EDFEEEEDEEEYFEFEFEFEFEFEFEFEFE 713
QY 1395 ASEVEK 1400
| :|||
Db 714 EGELEE 719

Search completed: November 6, 2001, 05:01:40
Job time: 7529 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 05:03:18 ; Search time 82.15 Seconds
(without alignments)
1309.294 Million cell updates/sec

Title: US-09-075-375A-6
Perfect score: 7495
Sequence: 1 MALSLHNVFLCKEALNLYA.....MEASEVEKLFKALPIRKVR 1412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2557	34.1	478	T03750	violaxanthin de-ep
2	2430	32.4	462	T00708	violaxanthin de-ep
3	210	2.8	522	A84606	hypothetical prote
4	194.5	2.6	2269	T28677	rhodopy protein -
5	167.5	2.2	2401	T28676	rhodopy protein -
6	158.5	2.1	1957	T38077	hypothetical colle
7	156	2.1	1365	T30822	lmp1 protein - Myc
8	155.5	2.1	886	T69378	conserved hypothet
9	155.5	2.1	2829	A42771	reticulocyte-bindi
10	155	2.1	2469	T36812	hypothetical prote
11	151	2.0	1939	T18372	repeat organellar
12	148.5	2.0	1935	A59286	myosin heavy chain
13	146	1.9	1935	A37102	myosin beta heavy
14	146	1.9	2748	S57976	nuclear migration
15	144.5	1.9	839	S54174	DNA topoisomerase
16	144	1.9	2166	T70163	hypothetical prote
17	143.5	1.9	1558	T71603	RESA-H3 antigen PF
18	142.5	1.9	880	T75103	conserved hypothet
19	142.5	1.9	1199	T29145	hypothetical prote
20	142.5	1.9	1819	T71928	cag island protein
21	142.5	1.9	1935	S06006	myosin beta heavy
22	142.5	1.9	5105	T32650	hypothetical prote
23	141	1.9	1002	T70319	nitrite reductase
24	140.5	1.9	1302	JC6009	surface-located me
25	140	1.9	800	T64508	hypothetical prote
26	139.5	1.9	1025	T55404	probable membrane
27	139	1.9	2712	T05113	hypothetical prote
28	137.5	1.8	1875	T38173	myosin-like protei
29	137	1.8	1109	A40801	phosphoprotein pho

30	137	1.8	1676	2	E71410	probable centromer
31	136.5	1.8	3724	2	T18427	hypothetical prote
32	136	1.8	1127	2	T28317	ORF MSV156 hypothe
33	135	1.8	3660	1	S02041	dystrophin, muscle
34	134.5	1.8	1937	2	T38055	myosin heavy chain
35	134	1.8	978	2	A70387	conserved hypothet
36	134	1.8	1639	2	S05603	major merozoite su
37	134	1.8	2261	2	T20978	hypothetical prote
38	133.5	1.8	1934	2	T48153	myosin heavy chain
39	133	1.8	1624	2	T25592	hypothetical prote
40	133	1.8	1979	2	C71622	hypothetical prote
41	132.5	1.8	1156	2	B70356	chromosome assembl
42	132	1.8	809	1	JQ0032	anthrax toxin leth
43	131.5	1.8	1170	2	A72287	hypothetical prote
44	131	1.7	1156	2	E69444	chromosome segrega
45	130.5	1.7	1631	1	SAZQK1	major merozoite su

ALIGNMENTS

RESULT 1

T03750
violaxanthin de-epoxidase precursor - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03750
R:Bugos, R.C.; Hieber, A.D.; Yamamoto, H.Y.
J. Biol. Chem. 273, 15321-15324, 1998
A:Title: Xanthophyll cycle enzymes are members of the lipocalin family, the first ide
A:Reference number: Z15054; MUID:98288256
A:Accession: T03750
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-478 <BUG>
A:Cross-references: EMBL:U34817; NID:gl463122; PIDN:AAC50031.1; PID:gl463123
A:Experimental source: strain Xanthi; tissue-type leaf
C:Genetics:
A:Gene: TVDEL
C:Function:
A:Description: violaxanthin de-epoxidase and zeaxanthin epoxidase catalyze the additi
d in protecting the photosynthetic apparatus from excessive light
A:Note: established as member of the lipocalin family
F:1-134/Domain: transit peptide (plastid) #status predicted <NPN>
F:135-478/Product: violaxanthin de-epoxidase #status predicted <NAT>

Query Match 34.1%; Score 2557; DB 2; Length 478;
Best Local Similarity 100.0%; Pred. No. 4.4e-142;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	474	MALAPHSNFLANHETIKYVVGSKLPGHKRFSWGWDYFGSIWVAKICSSRRIPRYFRKSP	533
DB	1	MALAPHSNFLANHETIKYVVGSKLPGHKRFSWGWDYFGSIWVAKICSSRRIPRYFRKSP	60
QY	534	RICCGLDSSGLQFLSHGKHNLSPAHSINQVPGNSGCKFPKDVLMYWEKGQAKTAI	593
DB	61	RICCGLDSSGLQFLSHGKHNLSPAHSINQVPGNSGCKFPKDVLMYWEKGQAKTAI	120
QY	594	VAIFILSVASKADAVDALKTCTCLLKECRLELAKCISNPACAANVACLOTNNRPDETEC	653
DB	121	VAIFILSVASKADAVDALKTCTCLLKECRLELAKCISNPACAANVACLOTNNRPDETEC	180
QY	654	QIKCGDLFNSVVDENECVSKKCVPRKSDVGFPPDPSPVLVQKFDPMKDFSGKWFIT	713
DB	181	QIKCGDLFNSVVDENECVSKKCVPRKSDVGFPPDPSPVLVQKFDPMKDFSGKWFIT	240
QY	714	RGLNPTDFADFCOLHEFHTEENKLVNLSWRITPDGGFFTRSAVQKFDVQDPKYGILYN	773
DB	241	RGLNPTDFADFCOLHEFHTEENKLVNLSWRITPDGGFFTRSAVQKFDVQDPKYGILYN	300
QY	774	HDNEYLLYQDDWYILSSKVENSPEDYIFYVYKGRNDAMDGYGGSVLYTRSAVLPESITPE	833

Db 301 HDNEYLIYQDDWYILSKSVENSPEDYIFVYKGRNDWDGGSGLVYTRSAVLPESTIPE 360

Qy 834 LQTAQKVGVRDNFTFIKTDNTCGPEPLVERLEKKVEEGERTIIKEVEIEEVEEKVRDK 893
|||||

Db 361 LQTAQKVGVRDNFTFIKTDNTCGPEPLVERLEKKVEEGERTIIKEVEIEEVEEKVRDK 420
|||||

Qy 894 EVTLFSKLFEGFKELQDEENFLRELSKEEMDVL DGLKMEATEVEKLFGRALPTKLL 950
|||||

Db 421 EVTLFSKLFEGFKELQDEENFLRELSKEEMDVL DGLKMEATEVEKLFGRALPTKLL 477
|||||

RESULT 2

T00708

violaxanthin de-epoxidase homolog F22013.3 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 22-Oct-1999

C:Accession: T00708

R:Shinn, P.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Sun, H.; Conway, A.; Con-
eologis, A.; Ecker, J.R.

A:Submitted to the EMBL Data Library, April 1998

A:Description: Genomic sequence for Arabidopsis thaliana BAC F22013.

A:Reference number: Z14200

A:Accession: T00708

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-462 <SHI>

A:Cross-references: EMBL:AC003981; NID:g3063438; PID:g3063441; GSPDB:GN00059; ATSP:F22013

A:Experimental source: Cultivar Columbia

C:Genetics:

A:Gene: ATSP:F22013.3

A:Map position: 1

A:Introns: 72/3; 128/2; 160/3; 292/2

hypothetical protein At2g21860 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84606
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
m.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: A84606
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-522 <SPO>
A:Cross-references: GB:AE002093; NID:g4417279; PIDN:AAD20404.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g21860
A:Map position: 2

Qy		304	YILSSOIQENKPDYIFVYVRGRNDAMDYGGSVIYTRS-----PTLPSIIPNLQ---KA	355
Dy		305	: : : : : : : : : : : : : : : : : : : : :	
Db		34	YTISNOIKTLN--VSTYPEGR-----EGTSSLELAKSWEKTLETITELTKSNBETVRL	87
Qy		356	AKSVGRDFNNFITDNSCGPEPPLVERLEKTAEBEKL-----LIKEAVEI	401
Dy		357	: : : : : : : : : : : : : : : : : : : : :	
Db		88	EKEIRELFKKY-----LDDEAEARKYLEGLUKLELNKKIKDIIAKIEYVNKTVEL	135
Qy		402	EEVEVK-----EVEKRDTEMTFORLLEGFKELQQ--DEENFVRELSKEEK	446
Dy		403	: : : : : : : : : : : : : : : : : : : : :	
Db		136	KKEIEKNNAVIDELANOSPKYVTGYTENKNTIYNTIKSPDYDIYEGDIDTFYNELSSI VK	195
Qy		447	E-ILNELOMBATEVEKFLGRALPI--RKURMALAPHNSFLANHET-----IKYIV	493
Dy		448	: : : : : : : : : : : : : : : : : : : : :	
Db		196	EDPIDDOIE-DKTKLENLRSKIDNVYDKIOKMJETVKSHLNETNKPNTILEIKKYI	254
Qy		494	GSKLPGHKRFSWGEDYFGSIVVAKICSSRRIPRYPRKSPRICGDLDSRGLQLFSH---	549
Dy		495	: : : : : : : : : : : : : : : : : : : : :	
Db		255	YDEI--SKELNKMLIEDFPNK-----EKELSNKISDYDKKREQUL--SEYKSKMLEIRHNYSQ	307
Qy		550	-----GKHNLSPAHSINONVP-----KGNSGCKFPKDV	577
Dy		551	: : : : : : : : : : : : : : : : : : : : :	
Db		308	TNYDNTKEEBEAKONYDKSNEHMTIPTNEDEISKIISEVTKMKDELTSKVNIYIDENK KY	367
Qy		578	ALMWYKWGFQAKTAIVAIPILVSASKADAVDAKTCCLLKCEURLELAKCISNPACAA N	637
Dy		579	: : : : : : : : : : : : : : : : : : : : :	
Db		368	KETVNSEHSOFTE-----LTDKIAEYSDK-----ELKK-----	396
Qy		638	VACLQTNCRNP--DETICOIKGDJFN-----SVVDENECAVSBRKCVPRKSD----	685
Dy		639	: : : : : : : : : : : : : : : : : : : : :	
Db		397	--CEQSFDNKNSLINETKNSIE--KEYQNTINTLUKKYVDEYIKVCCKSTESITAFSSKQTLL	452
Qy		686	-----VGDPFPDPSVLVQKFDMKDFSQKWFITRGLNPTTFDAFCQCLHEPHTEE-	734
Dy		687	: : : : : : : : : : : : : : : : : : : : :	
Db		453	KDMLNQNIKTVKETNSIDKS-YIEKFE-QILTCK--QTKLENKFTFESLINHEANNNEL	507
Qy		735	----NKLVGNLSWRITPDGGFTFRSAQKVFQDDPYPGILYNHDNEYLYLYQDDWTILSS	790
Dy		736	: : : : : : : : : : : : : : : : : : : : :	
Db		508	IKVFSDLKANLINGINEENMLYNQFTEK-EKTFNDIKEKNI--HINEEI-----S52	
Qy		791	KVENSPEDIYF-----VYYKGRNDAWD-----QY	814
Dy		792	: : : : : : : : : : : : : : : : : : : : :	
Db		553	KIEKIHASIYNISEETEREGINESLNTPKFVEKVENVTNUNKITEKLHYDFSDFGK	612
Qy		815	GGSVLYTRs-----AV-----LPESIPELOATAAQKVGREDENTFI	849
Dy		816	: : : : : : : : : : : : : : : : : : : : :	
Db		613	EGNKITYDKIKKINDDIMAYSQIDQHINGLDIOKKSSESYSVMKEQINKLEKVSNTETI	672
Qy		850	KTDNWCg---PEPPLVERLEKK--VEEGERTIIKEVEEEEE---VEKVRKDVTFLSKL	901
Dy		851	: : : : : : : : : : : : : : : : : : : : :	
Db		673	SNDNVEGIKKKOOTIYFKIDOKKKNIVBEIINKLSELSKTEKDKTSLSEYKDNLSYQNOL	732

Query Match

```

Db 1227 ILDIQISVEPFKI-----KFLKIKLURTKSSDCLKETDIEFKTISNLSDIQETKL-IEN 1280
QY 983 KRINGTFLKILKIPQISADLRITGGRSSRPLSAFRSGFSKGIIDFIVPLPSKNEKELTAP 1042
Db 1281 KNILNT-LEKLLLESLNQ-----KKNIED-----QKKELDEVNSK 1314
QY 1043 L-----LLKLVGLACAFLLVPSADAVDAKLTCAACLLKGCRIELAKCI 1085
Db 1315 IKNIESNVNQHKKNYEIGIVEXINEIAKA-----NKDQIESTQKLI-----IPTKLN 1362
QY 1086 ANPACAANVACLOTGNRPD-ETEQ-----IKGCDL-----FENSV 1121
Db 1363 ISPKANDLEGIDTNKLNKGYTEMNNIYEPIKSYDLITHYLETVSKPEITYEQINKR 1422
QY 1122 VDFNECAVSRKKCVPRKSDGLGEFPADPSVLQNP--NISDFNGKW--YITSGLNPTF 1176
Db 1423 ITAQNELLTNIKVNNAKSYLDIDIEANEDRITVTHFNKLNLDVNDKFTNEY--SKVNGKF 1480
QY 1177 DAFDCOLHEFHTEGDNKLVNLSWRILKTLDSGFFTRSAQVQVDPNPGVLYNHDEYEL 1236
Db 1481 DNTSINSINNKKSTDENLLNLINLOTKEM-----YANIVSKKYY-----SYK 1522
QY 1237 HYQDDWYLLSSKIENKPEYIFVYVYGRGRNDADWGGVAVVTRSPNSIPIPELKA 1296
Db 1523 YEAEINFINIPKAN-----SLNIQIKSSSGIDLFKNINI--AILPYLDSQKK 1568
QY 1297 SIGRDFSTIRTDNTCGPEPALVERIEKTIVERGI-----IVKEVEEIEEVEKE 1347
Db 1569 -----DTLTFI-----PSP-----EKTSETYTKISDSYNTLIDLKRSQELQKKQQA 1611
QY 1348 VEVGRTEMTLQRLAEGNFELKQDDENFVRLSEKEMEFLDEIKM---EASEVEKL 1401
Db 1612 LNLJ--FENRLLDKVKQATNELK---DTLSDLNKKEQILNKVKYLLHLKSNELNKL 1662

RESULT 6
T38077
hypoetical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38077
R:Connor, R.; Churcher, C.M.; Barrall, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z21767
A:Accession: T38077
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1957 <CON>
A:Cross-references: EMBL:Z70690; PIDN:CAA94624.1; GSPDB:GN00066; SPDB:SPAC1F3.06c
A:Experimental source: strain 972h; cosmid c1F3
C:Genetics:
A:Gene: SPDB:SPAC1F3.06c
A:Map position: 1

Query Match 2.1%; Score 158.5; DB 2; Length 1957;
Best Local Similarity 17.6%; Pred. No. 0.54;
Matches 265; Conservative 227; Mismatches 503; Indels 511; Gaps 68;

QY 59 YKTSFSDSHCKDKSQICSDTSTFEELQRLDKRGMTLLILEKQWRQFIQIAIVICTFV 118
Db 305 HNVATSDAIVHKDK-LIEDSTRISFEJNLKSERDTLSIKNEKLELLR----- 353
QY 119 IVPVDAVDALK--TCACLLKECRIELAKTIANPSCAANVACIQTGNRPDETECOIKC 175
Db 354 -----NTIGSLDRSTNSOLEEVEVLEK-----SN-----RTIHSQLTDAESKLSS 396
QY 176 GDLFENSV---VDQNECAVSRKKCVPRKSD-----VGEFPVDPDRNAVQNFN 220
Db 397 FEQENKSLKSGIDEYQNLLSSKDKMVKQVSQLPEARSSLAHATCKLAEINSEDFONKK 456
QY 221 MKDFSGKWIITSGLNPTDAFCQLHEFHM---ENDKLVLGNLTWRIKTLDG-GFFTRSAV 276

```

Db 457 IKDFE---KIEQDLRACLINSSNELKESALIDKDDQELNMLRQIQKQKVSSTQSSL 513
QY 277 QTFVQDPLPGALYNHNEFLHYQDDWYILSSQIENKPDYIFVYVRGRNDAMDGYGGSV 336
Db 514 QSLQD-----ILNEKKKHEVYESQLNELGELQTE-----544
QY 337 IYTRSTPESIIIPNLQAAKSVGRDFNNFITDSCGPEPPLVERLEKTAEEGEKLLIK 396
Db 545 -----INSHELSQLSTLAAEKAATNNE-----LSEKNSLOT 581
QY 397 EAVETEVEEVEKVRQTEM-----TLFORLLEGFKELQODEENFVRELSKEKEILN 450
Db 582 LCNAQERLAKSVQMLKENEQNFSSLDTSFKKLNESHQELNHHQITIKQL-KOTSSKLQ 640
QY 451 ELQMEATEVEK-----LFGRALPIRKLRLMALAPH-----SNFLANHETIKYVGS 495
Db 641 QLQLERANPEQESTLSDENNDLRTKLLKLEBSNKLKQEDVDSDLEKNIQTLK-----695
QY 496 KLPGRHKRFSGWEDYFGSIVVAKICSSRRIPRYFRKSPRICGLDSRGLQLF---SHGKH 552
Db 696 -----ED-----LRKSEALRFSK-----LEAKNLREVIDNLKGH 726
QY 553 NLSPA-----HST-----NONYPKNGSCGKPKPKVALMWKMGQFAKTAIVAFILSVA 602
Db 727 ETLEQRNDLHSSLDKAKNTNAILSELTKSSEDV-----KRLTANVETLTQD 774
QY 603 SKA-----DAVDALKTCCLKEKRELELAKCISNPACAAVACLOTNNRDETECOIK 656
Db 775 SKAMQSFTSLNYSQISNLHYELRHHVN-----MQSONNTLLESSEK 821
QY 657 --CGDLFEN--SWVDEF-----NECAVSRKVCVPRK-----SDVGDFFVPDP 694
Db 822 TDCENTQONTLIDNVQKLHMKHVNQESKSELKEVNGKLSLDLKNRSSLNVAISND 881
QY 695 SVLVQKEDM-KDFSGKWFITRGLNPTFDADFQCLHEFTTE-----NKLVCNLSWRIR 746
Db 882 QILTQLAELSKNYDSLEQESAQNSGLKSLAEKQLLHTTENEELHRLDKLTGKLGK----937
QY 747 TPDGFFTSRQVQKFDQPKYIGILYNHNEVLLYQDDWYILSSKVENSPEDYIFVYKG 806
Db 938 -----TEESKSDL-----GKLTARQEE-----ISNLKENNSQATSVKS 976
QY 807 RNDAMDGYGGSVLYTSRSLPESIIPELQTAQKVGCRDFNTIKTDTCGPEPPLVERLE 866
Db 977 KLDE-----TLSSSKL-RADIEHLKKNVSEVEVERNALLASN-----ERLM 1017
QY 867 KVE-EGERTIIEVEEIEEVEKVRDEKVTILFSKL-----FEGF-----KELQD 911
Db 1018 DDLKNGE-----NIASLQTEIEKKRAENDDLQSKLSVVSSEVENLLLTSSQTNKSLE-D 1071
QY 912 EENFLRELKSEMDVLDGLKMEATEVEKL--FGR-----ALPIRKLMAVATHC 957
Db 1072 KTNQLYIEKNVQKLLDEKQDRVLEELTSKYKLGEEAQAQIKDELLALRKK-----1124
QY 958 FTSPCHDIRFFSSD-----DGIGRL-----GITKRLNGTFLKILPIQS--AD 1001
Db 1125 -SKQHDLCANFVDDLKESDALEQITNEKNELIVSLEOSNNALVEERSDLANRLSD 1183
QY 1002 LRTTGRSSRPLSAFRSGFSKGFIDIVPLPSKNELKELTAPLLKLVGLVACAFILVPSA 1061
Db 1184 MKKSLSDSDNVISVIRS-----LVRVN 1206
QY 1062 DAVDALKTACILLKGGRIELAKCIANPACAAVACLOTNNRDETECOIKCGDLFENS 1121
Db 1207 DELDTLK-----RKDSLSQYSEVCQDRDLDSLKGC-----1240
QY 1122 VDEFNECAVS-RKKCVPRKSDLGEFPAPDPSVLQVN--FNISDFNGKWIT-SGLNPTD 1177
Db 1241 EESFNKAVSLRELCTKSEIDV-----PVSEILDDNFVNAGNFSLSRLTVLSLENYLD 1295
QY 1178 AFDCQLHEFTEGDKLVGNISWRIKTLDSGFFTSRAVQKFDQNPQGVLYNHDEYHL 1237
Db 1296 AFN-QVNFKKWELDN-----RLITTDADF-----TKVAVDL-----EKLQH 1330

QY 1238 YQDDWYILSSKIENKPDYIFVYVRGRNDAMDGYGGAUVYTRSSVLPNSIPIPELKAAS 1297
Db 1331 EHHDWLI-----QRG-----DLERAKLD 1348
QY 1298 IGRDFSTFIRTDNTCGPEPALVERIEKTVVEEGRIRIIVEVEEIEEVE-----1345
Db 1349 SEKNF-----LRKEAEMTENIH-SLEEGKEETKKEIAELSSRLDNLQATNKLKNQ 1398
QY 1346 -----KEV-----EKVGRTEMFLFORLAEFNGELQODEENFVRELSKEKEEFL--DEIKMEA 1395
Db 1399 LDHLNQELIRKEDVLKESLIISLESLSNQOKESSLLD--AKNEHMLDDTSTRKNS 1456
QY 1396 SEVEKL 1401
Db 1457 SLMEKI 1462
RESULT 7
T30822
Impl protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 07-Dec-1999
C:Accession: T30822
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christiansen, G.
Infect. Immun. 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis PG21 deletion mutants by cultivation in the
A:Reference number: Z18884; MUID:95369882
A:Accession: T30822
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1365 <JEN>
A:Cross-references: EMBL:U21962; NID:g790243; PID:g790244; PIDN:AAA81013.1
C:Genetics:
A:Gene: Impl
A:Genetic code: SGC3
Query Match 2.18; Score 156; DB 2; Length 1365;
Best Local Similarity 16.8%; Pred. No. 0.46;
Matches 209; Conservative 187; Mismatches 420; Indels 430; Gaps 52;
QY 272 TRSAVQTFVQDPLPGALYNHNEFLHYQDDWYILSSQIENKPDYIFVYVRGRNDAMD 331
Db 282 TRNQIQEINT-----NKNNP--NYSE-----LISQTSKRD-----SKNSVTD 319
QY 332 YGGSVIYTRSTPLPESII-PNLQKA-----AKSVGRDFNFIITDSCGPEPPLVERLE 385
Db 320 SNKSDIESANTELQALAKANADKQADNLAKISKEQLNNSVSNANT-----LSA 369
QY 386 TAECEKELLKEAVEIEEVEEVEKVRDEKVTILFORLLEGF-----KELOQDEENFVREL 441
Db 370 KLTDKNTIQQAKTELEKEVQKADQAIKSNNTASMQSAKSSLDKAEITKKLETF-----425
QY 442 SKEEKEILNQMEATEVEKLFGRALPIRKLRLMALAPHSNFLANHETIKYVYVSKLPQKH 501
Db 426 NKDKEAKFNLKQTRNQIQEF-----446
QY 502 RPSGWEDYFGSIVVAKICSSRRIPRYFRKSPRICGLDSRGLQLFSGHKHNLSPAHS-I 560
Db 447 -----INTKNPNYSELTSQTSKRSKNSVTDSSNKSIDIESANTEL 489
QY 561 NONVPKNGSGCKFPKPDVALMWKMGQFAKTAIVAFILSVASKADAVDALKTCTCLLKE 620
Db 490 KOALAKAN-----ADKVOA-----DNLAAS 509
QY 621 CRLELAKCISNPACAAVACLOTNNRDETECOIKCGDLFENSVDDEFNECAVSRKCV 680
Db 510 IKEQLNNSVSN-----ANTLSAKLTDRDN--TIQQAkte 541
QY 681 PRKSDVGFPPDPVSVLVQKMDKDFSGKWFITRGLNPTFDA-----FDCQLHEFTEENK 736
Db 542 LEKE-----VQKADQAIKSNNTASMQSAKSSLDKAEITKKLETFNKDKEA 588

Query Match 2.1%; Score 155; DB 2; Length 2469;
Best Local Similarity 17.3%; Pred. No. 1.2;
Matches 230; Conservative 208; Mismatches 497; Indels 399

QY	271	FTRS	AVOTFVQD	PDLFCALYNHNEFLHYQDDWYILSSOIQENKPDYIFVYR-----	323
D6	178	FSTS	NVDAIEYLSPPNVOYT--GSELYFPKEYIGHSH-----YTMNHYRVINYEKL	228	
QY	324	-GRD	AMDGXGGSVIV-----TRSPITPESIIPLNQAKAASVGDRDNFI	367	
D6	229	HGPN	IDLTSGEGLIELSPNTPKPTSTOKPPTRTPKPATPKAPKTPRPKPKESTI	288	
QY	368	TTDN	SCGPPEPLVERLEKAEBEGEKLLIKEAV-----EIEBEVEKEVKYVDT--EMTL	419	
D6	289	PYDS	--KKPP--KIPKTSKSKVLTKDTALTTPQHKTIBEHLRELLPPIETEVEONTL	343	
QY	420	FOR	-----LLEGFEKLQ-----QDRENFRVLSKEEKEILNELOMEATEVEKL	462	
D6	344	FNH	VERTTGTDSLISLGINSITTKREDDEDDONTWSKLKEDEDWDID-----	392	
QY	463	FGRA	LPIRKIRLMALAPHSNFLANHETIKYYGSKLPGHKRFSGWEDYFGSIVVA-----	517	
D6	393	---IP	TEVLDTFETHS---DOETI-YMIGDE--NIHDWSYSDDDDIDDTLDSFIQLD	441	
QY	518	-KIC	SRRIPRYFKSPCCGLDSRGLOLFSSHGK-----HNLSPAHSINQNVPKG	567	
D6	442	NLTS	LDNIPKN-NTFPRJ---IDKTSNOPIKEGHALHSIDRIKLNVILEHGLITSISIS	497	
QY	568	NSCK	PFKDVALMWBKWGQFAKTAIVAIFILSVASKADADVAKTCCLKECRLEIAK	627	
D6	498	ISKCS	LLQFVIL--WGE-----KLSIPTR-DLKTKLTTELIIETAEIALTK	542	
QY	628	CISN	PACAAVACLQTCNNRPDETQCICGDLFENSUVDEFECASVRKKCVPRKSDVG	687	
D6	543	-LTN	-----DTRFNNVITKLNCMKLKL-----SFVSVD	569	
QY	688	DFF	-----VPDPSSLVQKFDM-----KDFS-----GKWEIFTRLGNLPTF	720	
D6	570	SYKH	SALLNLIKIQTIDTELKTLTVTSELGXDFSVCTKKESETIMAAINKLK	628	
QY	721	DAFD	COLHEFHTEENKLVGNLSWRHTPDGFFTSRSAVOKF-----VODKYPG	769	
D6	630	EKSTR	QELTHEENTFQSVL-----IAMETQPILPTRVIETIOFPSKKAQ	675	
QY	770	ILYN	HDNEVILLYQDDWYILSSKVENSEPDIYPVYKGRND---AWDGYGGSVLVTRSAVL	826	
D6	676	QL--	HEKS	KLVEQK----LTIDANNVLTDLLHTMKQDKTDISPAPD-----FTTVLKN	722
QY	827	PESI	IEPLOTAAQKVCRD---FNPIKTDNTCGPPEPLVERLEKKVEBERTI-IKEVEE	882	
D6	723	IQSL	QLLOQTCTVDUNIDKKFISNTVQOOLSYIGWEVAELSOSHQWNFPKADEVIPKLTI	782	
QY	883	IEEE	BEKVRDKEVTLFSKLFEFGKELQRDEENFLRELSEMDVIDGLKMEATEVEKLF	942	
D6	783	IKKEI	OQVTTKQ-----KNEETLSKILADVOTLLENQAQSDTSLPILOHVITKAGTLVG	837	
QY	943	RALP	IRKLMATHCFTSCHDRIRFRFSDDGIGRLGIT-----RKRINGTPLLLKILPPIQ	998	
D6	838	E-----	RENQFESKNTVQKLSTSEEFCLKTIDSTLENVQLIQ	878	
QY	999	S-ADR	LTTGGRSRPLSAPRSFGK---IFDIVPLPSKNELKELTAPLLL-----	1045	
D6	879	EISD	ILSQNOXIHOSETIKAFFDKSNTIINNILQINOQXYTTVTOPMLIANKRFISEA	938	
QY	1046	-----	KLVGVLCAPLIVPSADAVDALKTCACLLKGCRIELACIANPACAANVAC	1096	
D6	939	KFRE	SNTICEIISTISLUGSLSKSTTYEALKDKALKSIDTLKEKL-----	983	
QY	1097	LOTC	NNRPDETE-----CQIKCGDLFENSVDENBCASVRKKCVPRKSDLGERPAD	1149	
D6	984	--TAV	DRPLKRELYNVIRKLOKLTLEQQEFFDW-----KNEVDSF-VPT	1027	
QY	1150	PSVL	VONF--NLISDFNGKWYITSGNLNPFTFADPCQLHEHFTPEGDN-KLVGNISWRKITLD	1206	
D6	1028	PSRO	KTFEQNAFPMKAKOYAKALKDQIOAMEIDVDPESVIEDNIKANGQAKWO----	1082	

```

QY 1207 SGFFTRSAVQKVPQDPNGPGLVLYNHDNEYLHVQDDWYILSSKIENKPEDYIF-----VYY 1261
      :||| ||| :: ||| |::| ::| :|
Db 1083 -----RQSFAQDNLNFSLI-----PDWLSL-AKEYTRPKSTLFTVGIPILL 1124
      :||| ||| :: ||| |::| ::| :|
QY 1262 RGRNDADWDGYGAVVYTRSSVLPSNII-----PELEKAAK 1296
      :||| ||| :: ||| |::| ::| :|
Db 1125 KEVEEVLSEVKLNKEAKLKSLLPNGPVTPPPKFDMHIHYESNVNPHLKTINLPKVSTVAH 1184
      :||| ||| :: ||| |::| ::| :|
QY 1297 SIGRDFSTFIRTDNT-CGPEPALVERIE-----KTVE-----EGERIIVKEV 1337
      :||| ||| :: ||| |::| ::| :|
Db 1185 NIGHELSSLLOALNSKTLPEAVGVGTSLBQHAAKFSCKMEKLTLEATWHDHQVDRTKIDEVI 1244
      :||| ||| :: ||| |::| ::| :|
QY 1338 BEIEBEVEK-----EVEKVR-----TEMTLFORLAGFNELKQDEENVRELSKEMEFEL 1388
      :||| ||| :: ||| |::| ::| :|
Db 1245 EDLRNDTKHHVAPQIQSPNRFLLSPEDIOETNSLPKFLRDSLLENESRLLASOKNEFOML 1304
      :||| ||| :: ||| |::| ::| :|
QY 1389 DEIKMEASEVE 1399
      _::|_|_|_|
Db 1305 -ENTVKAARELQ 1314
      _::|_|_|_|

RESULT 11
Ti8372
repeat organellar protein - Plasmodium chabaudi
C:Species: Plasmodium chabaudi
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: Tl8372
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
A:Title: A plasmodium chabaudi protein contains a repetitive region with a predicted
A:Reference number: Z18922; MUID:98418765
A:Accession: Tl8372
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1939 <WR>
A:Cross-references: EMBL:U043145; NID:g1151157; PID:g1151158; PIDN:AAC63403.1

```

Query Match	2.0%;	Score 151;	DB 2;	Length 1939;
Best Local Similarity	17.5%;	Pred. No. 1.5;		
Matches 189;	Conservative 167;	Mismatches 299;	Indels 422;	Gaps 50;

```
QY 847 TFIKTDNTGCPPEPLVERLEKKEVEGERTIIKEVEIEEEVKVR-----DKEVT 896
Db 414 LEIE-----TIKNELEKEEKELEKNEKKAHVVTNTITKEIK 449
QY 897 LFSKLPFGKELQDE-----ENFLRELSKEEMDVLGLMEATEVEKLFGRALPIRKLMA 952
Db 450 LLEKKTKEGKHNELENNLSKNEK-----DNIRKNETEL-----490
QY 953 VATHCFTSPCHDIRFSSDDGIGRIGITRKIRINGFTLLKILPPIOSADLRTTGGSSRP 1012
Db 491 -----NDKISSLNSE-----VN-----ILNKDKQTIG-----512
QY 1013 LSAPRSFSGIGFDIVPLPSKKNELKELTAPLLLLKLVGLACAFIIVPSADAVALKTCAC 1072
Db 513 -----NDIKTLN-----DLINLNK-----527
QY 1073 LLKGCRIELAKCIANPACAAVACLQTCNNRPDE-----TEQIKCGDLFENSVD 1123
Db 528 -----EINTSDKNKMKEDLAMLNEEGKC-----VVID 558
QY 1124 EF-----NECAVSRKKCVPRK-----SDLGEFPAPDPSPVLVQNFNISDPNGKWYITSGLNPTF 1176
Db 559 EIEKKYKNEIFMLELKEKENVADLND-----EISILRNSIYVKE-----KEFI-----603
QY 1177 DAFDCQLHEFHTGDNKLVGNISWRITLDSGFFTSVAVOK-----FVQDPNQCVLYNHND 1233
Db 604 -----EMKEFY-----ENKI-----NLFNKNFEKKNIYENELNLSURLKYDNEQ 642
QY 1234 EYLHQDDWILSKTENKPEDYIFVYGRNDADWDGYGAVYVTRSSVLPNSIIPLEK 1293
Db 643 GLIKQIDELNIQKLEEK-----YLQY-----NDN-----MHMERSICTKIDMYPSEN 687
QY 1294 AAKSIGRDFST-FI--RTDNTGCPPEPALVERIEKTVVEGERIIVKEVEIEEVEKEVEK 1350
Db 688 IKGSDLVDFVTAYIKRDESSDANPDTHK--EMVAELEKRAHAAIVAELEEKHKEEIAK 745
QY 1351 VGRTEMFLFORLAEFNE-----LKQDENFVRELSEMEFDEIKMEASEVEKLP 1403
Db 746 LBGHKEVVLRLGEHQKEETIIILEEKHKDVVTKLGEQHKENIILKEEHKDVVTKLG 802
```

RESULT 12

```
A59286
myosin heavy chain beta chain, cardiac - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C:Accession: A59286
R:Ko, Y.L.
submitted to GenBank, October 1996
A:Reference number: A59286
A:Accession: A59286
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1935 <KO>
A:Cross-references: GB:U75316; NID:g1698894; PIDN:AAB37320.1; PID:g1698895
A:Experimental source: strain domestica
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: cardiac muscle; heart
F:88-766/Domain: myosin motor domain homology <MMO>
```

```
Query Match 2.0%; Score 148.5; DB 2; Length 1935;
Best Local Similarity 16.7%; Pred. No. 2;
Matches 248; Conservative 220; Mismatches 555; Indels 459; Gaps 57;

QY 166 PDETECQIKCGDLFENSVDQFNECAVSRKKCVPRKSDVGEFVPPDRN-----AVV 216
Db 375 PDGTE-----EADKSAYLMLGNSADLLKGLCHPRKVGNEYTKGQNVQVYATGALA 428
QY 217 QNFNKNDFSGKWYITSGINPT-----FADFDCQLHE-----FHMNDKL 255
Db 429 KAVYEKMFN--WMYTR--INTTLETKQPRQVFIGVLDTAGFEIFDENSFEQLCINFNTNEKL 485
```

```
QY 256 -----VGNLTWRIKTLTDGGFTTSR SAVOTE-----V 280
Db 486 QOQFNHMFVLEQEBYKEGIEWEF--IDFGNDLOACIDILIEKPMGIMSILIEECMPKA 543
QY 281 QDPDLPGALY--NHDNEFLHYQDDWYILSSQIENKPD--DYIFVYGRND-----AMDYGG 334
Db 544 TDMTFKAKLYDNHLGKSNFQK-----PRNIKGRPEAHFALIHVAGTVDYNIIGW-----593
QY 335 SVIYTRSTPLPESIIPIINQKAA--KSVGRDFNNFITDNSCGPEPPLVERLEKTAEEBKL 393
Db 594 --LQNKDPLNETVVDLYKSKSLKLLSNLFANY-----AGADTP--VEKGRKAKKGSF 644
QY 394 LIKAEVETEEVEKEVEKVRDENTLFLORLLEGFKELQDDEENFVRELSKEKE-----447
Db 645 QTVSALH--RENLNKMTNLRST-----HPHFVRCIIPNETKSPGVID 685
QY 448 ---ILNELQMEAT--EVEKLFGRALP-----IRKLMALAPHSNFLANHETIKY 492
Db 686 NPLVMHQLRNCVGLBGIRICRKGPNNRILYGDFRQRYRILNPAAIPEGQFIDSRKGAEL 745
QY 493 VGSKLPGHKRPSWGWEDYFGSIVVAKICSSRIIPRYFKSPRICGGLDSRGLQLFSHGKH 552
Db 746 LGSLLDIDHNOYKFGHTKVFFKAGLGLLEEMRDERLSRIITRI--QAOSRGVLSRMEFK 803
QY 553 NLSPAHS---INQNVPKNGSGCKFPKDVALMVWEKMGOFAKTAIVAIFILSVASKADAVD 609
Db 804 LLERRDLSLLIOWNIR-----AFMSVKNW--PWNKLYFKIKPLLESATEKEMA 850
QY 610 ALKTCCLLKEC-----RLELAKCISNPACAAVACLQTCNNRPDETEQIKCGDLFE 662
Db 851 TKKEFGRLKEALEKSEARRKELEKMYSLLEKNDLQVQAEQDNLSDSEERCDQLIK 910
QY 663 NSV-----VDENECAVSRKK-----CVPRKSDVGDVFPVDPSPVLVQKFD 702
Db 911 NKIQLEAKVKEMTERLEDEEEMNAELTAKRNVEDCESELAKRIDDLLETLTAKVEKEK--968
QY 703 MKDFSGKFWITGLNPTDFADFCQLHEFHTENKLVGNLSWRIRTPDGGFTTSR SAVOKFV 762
Db 969 -----HATENK--VKNLTEEMAGL-----ETIAKLT 993
QY 763 QDPKPGYILYHNHNEVLLYQDDWYILSSKVENSPEDYIFVYVYKGRNDADWDGYSVLYTR 822
Db 994 KEKALQEAHQALDDLOAEEDKVNTLTAKVKLQHV-----DDLGSLEQEK 1042
QY 823 SA-----VLPEIIPELOTAQKVG-----RDF-----NTFIKTONTGCP 857
Db 1043 KYRMDLERAKRKLEGLDLKLTQESIMDLENDKQOOLDERLKKKDFELNALNARIEDSQALGS 1102
QY 858 E-----PPLVERLEKKEVEGERT-----LIKEVEIEEVEKEK-----889
Db 1103 QLOKKLKELOARIEELEEL--EAERTAKRAKVEKLSRDSLELEISERLEAGGATSVQI 1161
QY 890 -----VRDKEVTLFSLKLFEGFKELQDDEENFLREL 919
Db 1162 ENMKKREAEFFQWRDRLEATLQHEATAAALRKKHADSVAGELGEQIDNLQVKQKLEK 1221
QY 920 SKEEMDVLGLMEATEVEKLFGRALPIRKLMVATHCFTSPCHDIRFSSDDGIGRIG 979
Db 1222 SEFKLELDD---VTSNMEQIIKAKANLEKM-----CRT--LEDOMNEHRS-----KAE 1264
QY 980 ITRKRINGFTLLKILPPIQSADLRTTGGSSRP---SAFRSGFGSKGIFDIVPLPSKNEL 1036
Db 1265 EFORSVNDL-----TSORAKLOTENGELSRLQDEKALISQLTRG-----KLTYTQOL 1312
QY 1037 KELTAPLLKLVGLACAFIIVPSADAVALKTCACLLKGCRIELAKCIANPACAAVAC 1096
Db 1313 EDLKRQLEEEVKAKNALAHALQASARHAADLLREQYEEETETKAELOQVLISK--ANSEVAQ 1370
QY 1097 LOTCNNRPDETECQIKCGDLFENSVDNECAVSRKKCVPRKSDLGEFPAPDPSPVLVQ 1156
Db 1371 WRT-----KYETDAIQRTTEELEEAKKLAQRLQDAEE-----1402
```

QY 1157 FNISDFNGKWIYISGLNPTDFAFCOLHEFHETBGDNKKNLVGNISWRIKTTLDSDGFFTRSAVO 1216
 Db 1403 -----AVEAYNAKSSLEKTKHRLQNEIDLMVD-----VERSNA 1438
 QY 1217 KFQDPNPQGVLYNHNDYILSKYKLPEDYIFVYGRNDANDGYGAVV 1276
 Db 1439 AALDKKQ-----RNFKILAEWKQKVEESQESKE-----ARSL 1477
 QY 1277 YTRSSVLPNSIIPELE-----KAAKSIGRDFSTFRTDNTGCGPEPALVERIEKTVEEGE 1330
 Db 1478 STEFLKLNAYEESLEHLETSKREKNLQEEISDLTEQLGSSGKTHLEKVKRQLEAEK 1537
 QY 1331 RIIVKVEETEEVEKEVKGVTETLQ-----RLAEGFNEUKQDEENFVRELS-- 1381
 Db 1538 LEQSALESEASLEHDEGKILRAQLEFQKAEKMERKLAEDMEQAKRNLHRYVDSL 1597
 QY 1382 -----KEEME-FLDEIKMEASEVEKLFCKA 1405
 Db 1598 QTSILDAETRSRNEALRVKKMEGDLNEMEIQLSHANPMAEA 1639
 RESULT 13
 myosin beta heavy chain, cardiac and skeletal muscle - human
 N:Contains: myosin ATPase (EC 3.6.1.32)
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Jan-2001
 C:Accession: A37102; S12733; A94224; B28908; A24997; A27858; I54254; S12458; S09331; S02
 R:Jaenicke, T.; Diederich, K.W.; Haas, W.; Schleich, J.; Lichter, P.; Pfordt, M.; Bach,
 Genomics 8, 194-206, 1990
 A:Title: The complete sequence of the human beta-myosin heavy chain gene and a comparati
 A:Reference number: A37102; MUID:9105634
 A:Accession: A37102
 A:Molecule type: DNA
 A:Residues: 1-1935 <JAE>
 A:Cross-references: GB:M57965; GB:M30603; NID:g179507; PIDN:AAA51837.1; PID:g179508; GB:
 R:Liew, C.C.; Sole, M.J.; Yamauchi-Takihara, K.; Kellam, B.; Anderson, D.H.; Lin, L.; Li
 Nucleic Acids Res. 18, 3647-3651, 1990
 A:Title: Complete sequence and organization of the human cardiac beta-myosin heavy chain
 A:Reference number: S12733; MUID:90301496
 A:Accession: S12733
 A:Molecule type: DNA
 A:Residues: 1-106, 'E', 108-177, 'E', 108-671, 'LVH', 675-857, 'A', 859-941, 'NV', 944-1123, 'A', 1125-1158, 'C',
 A:Cross-references: EMBL:X51591; NID:g29726; PIDN:CAA37068.1; PID:g29727
 R:Yamauchi-Takihara, K.; Sole, M.J.; Liew, J.; Ing, D.; Liew, C.C.
 Proc. Natl. Acad. Sci. U.S.A. 86, 3504-3508, 1989
 A:Title: Characterization of human cardiac myosin heavy chain genes.
 A:Reference number: A94224; MUID:89264452
 A:Accession: A94224
 A:Molecule type: DNA
 A:Residues: 1-87, 'Q', 89-106, 'E', 108-177, 1325-1702, 'DR', 1705-1786, 1788-1803, 'E', 1804-1935
 R:Yamauchi-Takihara, K.; Sole, M.J.; Liew, J.; Ing, D.; Liew, C.C.
 Proc. Natl. Acad. Sci. U.S.A. 86, 7416-7417, 1989
 A:Reference number: A94226
 A:Contents: annotation; erratum
 R:Kurabayashi, M.; Tsuchimochi, H.; Komuro, I.; Takaku, F.; Yazaki, Y.
 J. Clin. Invest. 82, 524-531, 1988
 A:Title: Molecular cloning and characterization of human cardiac alpha- and beta-form my
 human atrium.
 A:Reference number: A92770; MUID:88299163
 A:Accession: B28908
 A:Molecule type: mRNA
 A:Residues: 1412-1518, 'R', 1520-1574, 'NV', 1577-1935 <KUR>
 A:Cross-references: GB:M21665
 A:Note: the authors translated the codon AGC for residue 108 as Arg
 R:Lichter, P.; Umeda, P.K.; Levin, J.E.; Vosberg, H.P.
 Eur. J. Biochem. 160, 419-426, 1986
 A:Title: Partial characterization of the human beta-myosin heavy-chain gene which is exp
 A:Reference number: A24997; MUID:87030293
 A:Accession: A24997
 A:Molecule type: DNA
 A:Residues: 682-721; 975-1112; 1854-1935 <LIC>
 A:Cross-references: GB:X04627

R:Saez, L.J.; Gianola, K.M.; McNally, E.M.; Fedhali, R.; Eddy, R.; Shows, T.B.; Leinw
 Nucleic Acids Res. 15, 5443-5459, 1987
 A:Title: Human cardiac myosin heavy chain genes and their linkage in the genome.
 A:Reference number: A93669; MUID:87260010
 A:Accession: A27858
 A:Molecule type: DNA
 A:Residues: 1854-1865, 'A', 1867-1935 <SAE>
 A:Cross-references: GB:X05631; GB:Y00362; NID:g34643; PIDN:CAA29119.1; PID:g34644
 R:Diederich, K.W.; Eisele, I.; Ried, T.; Jaenicke, T.; Lichter, P.; Vosberg, H.P.
 Hum. Genet. 81, 214-220, 1989
 A:Title: Isolation and characterization of the complete human beta-myosin heavy chain
 A:Reference number: I54254; MUID:89154425
 A:Accession: I54254
 A:Status: translated from GB/EMBL/DDEJ
 A:Molecule type: DNA
 A:Residues: 653-720 <RES>
 A:Cross-references: GB:M27636; NID:g179511; PIDN:AAA79019.1; PID:g601916
 R:Bober, E.
 submitted to the EMBL Data Library, January 1989
 A:Reference number: S12458
 A:Accession: S12458
 A:Molecule type: mRNA
 A:Residues: 785-1076, 'E', 1078-1123, 'A', 1125-1702, 'DE', 1705-1935 <BOB>
 A:Cross-references: EMBL:X51591; NID:g29467; PIDN:CAA35940.1; PID:g29468
 R:Bober, E.; Buchberger-Seidl, A.; Braun, T.; Singh, S.; Goedde, H.W.; Arnold, H.H.
 Eur. J. Biochem. 189, 55-65, 1990
 A:Title: Identification of three developmentally controlled isoforms of human myosin
 A:Reference number: S09331; MUID:90235862
 A:Accession: S09331
 A:Molecule type: mRNA
 A:Residues: 785-830, 'X', 832-900, 'X', 902-970, 'X', 972-1040, 'X', 1042-1076, 'E', 1078-1110,
 '1602-1670, 'X', 1672-1702, 'DE', 1705-1740, 'X', 1742-1810, 'X', 1812-1935 <BOV>
 R:Jandreski, M.A.; Liew, C.C.
 Hum. Genet. 76, 47-53, 1987
 A:Cross-references: EMBL:X51591
 A:Title: Construction of a human ventricular cDNA library and characterization of a b
 A:Reference number: S02229; MUID:87192738
 A:Accession: S02229
 A:Molecule type: mRNA
 A:Residues: 1393-1702, 'DR', 1705-1935 <JAN>
 A:Cross-references: EMBL:X06976; NID:g34860; PIDN:CAA30039.1; PID:g825694
 R:Saez, L.; Leinwand, L.A.
 Nucleic Acids Res. 14, 2951-2969, 1986
 A:Title: Characterization of diverse forms of myosin heavy chain expressed in adult h
 A:Reference number: A93616; MUID:86176778
 A:Accession: B23767
 A:Molecule type: mRNA
 A:Residues: 'LLGVGELASG', 1311-1312, 'G', 1314-1355, 'R', 1357-1358, 'GD', 1361-1438, 'LQ', 14
 A:Note: the first ten codons of the sequence figure show the reverse complementary st
 C:Genetics:
 A:Gene: GDB:MYH7
 A:Cross-references: GDB:I20215; OMIM:160760
 A:Map position: 14q12-14q12
 A:Introns: 67/3; 115/3; 168/1; 177/2; 213/3; 244/3; 266/1; 299/1; 333/3; 380/1; 419/3
 24/3; 1390/2; 1451/3; 1507/1; 1548/3; 1651/3; 1719/3; 1761/3; 1853/3; 1885/3; 1930/3
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; hydrolase; methyl
 F:88-766/Domain: myosin motor domain homology <MMOT>
 F:178-185/Region: nucleotide-binding motif A (P-loop)
 F:548-585/Region: actin binding #status predicted
 F:655-677/Region: actin binding #status predicted
 F:839-1935/Domain: coiled coil #status predicted <COI>
 F:839-1279/Region: S2
 F:1280-1935/Region: light meromyosin
 F:129/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
 F:184/Binding site: ATP (Lys) #status predicted
 F:695,705/Active site: Cys #status predicted
 Query Match 1.9%; Score 146; DB 1; Length 1935;
 Best Local Similarity 16.9%; Pred. NO. 2.9;
 Matches 238; Conservative 207; Mismatches 529; Indels 434; Gaps 51;

QY 276 VQTEVQDPLPGALYNHNDNEFLHYO-----DDWYILSQIENKPDYIFVYIRG 324
Db 949 YE-----DLVKCKENPMFLKESAKLGHVTVVSNAYSELEKKLEQPSLEYLVHAKA 1002
QY 325 RND---AWDGGGVS-----IVRSPITPESIPN-----LOKAASKVGRDF--- 363
Db 1003 TNHLLSDSAYEDLVCKENPMDFLEKESAKLGHVTVVSNAYSELEKKLEQPSLEYLV 1062
QY 364 ---NFIITDNS-----CGPEPPLVERLEKTAEGEKKLLKEAV-ELEEEVEKE--- 408
Db 1063 HAKATNHLLSDSAYELVCKENPMDFLEKESAKLGHVTVVSNAYSELEKKLEQPSLEY 1122
QY 409 --VEKRDTEMTLFORLEGEKELQDEENFVRELKKEKEIINE--LMEEA-TEVEKLF 463
Db 1123 YLVHAKATNHLLSD--SAYELVCKENPMDFLEKESAKLGHVTVVSNAYSELEK-- 1178
QY 464 GRALPIRKLWALAPHNSFNLANHETIKYVGVSKLPGHKRFSWGWDYFSGIVVAK 518
Db 1179 -----KLEO---PSLAYLVEH-----AKATDHLLS---DSAYEDLVCKENPDV 1217
QY 519 -----ICSSRRIPRYPRKSPICCGDLSRGL-OLFSHGKHNLSNPAHSINQNV 564
Db 1218 EFLKEKSAGLGHVTVVSNAYSELEK-----LEQPSLAYLVEHAK--ATDHLLSDSA 1268
QY 565 PKNSGCKFPDVALMVWKEWQGFAPTAIVAIFILSVASKAD-----AVDALKTKTCL 617
Db 1269 YEDLVCKENPMDFLEK--EKSAGLGHVTVVSNAYSELEKKLEQPSLEYLVHAKATNHLL 1327
QY 618 LKECRLE-LAKCISNPAC-----AANVACLOTQNNRPDETEQIKCGDLFENSVDEN 670
Db 1328 LSDSAYEDLVCKENPMDFLEKESAKLGHVTVVSNAYSELEKKLEQPSL----- 1377
QY 671 ECASVRKKCVPRK--SDVGDF-----PVPDPSVLVQKDFMDKDFSGKWFITRGLNPTDFAF 723
Db 1378 EYLKHAEOIQSKIISDNTLANPSMEDMASKLOKLEYQIVSNDYI--ALKNTMEKP 1435
QY 724 DCQ-----LHEEH-----TEENKLVGNLSWRITPDGGFTFSRAVQFVOD--PKYPGILYN 773
Db 1436 DVELLRSKLGHIIDTTTVELVSNF-----SPTLFIBEKAKSGYRLI 1482
QY 774 HDNEYL-----LYQDDWYILSK-----VENSPE-----D 798
Db 1483 EPNEYLDLNRITATPSKEEDNFCQIGCVALDSKEVERLKNLSNPKSKFFTEENAALLD 1542
QY 799 YIFV---YYGRNDAMDYGGSVLYTRSAVLPSIPELOTAQAKVGRDPNFTIKDNTC 855
Db 1543 LVLVDKTEYOAMD-----NASNKSLLIPSTKAL-----DFVTM----- 1576
QY 856 GPPEPLVERLEKKEVEGERTIIEVEEIEBEVEKVRDKEVTFLSKLFEKGLQORD--- 911
Db 1577 -PAPOLASAEKSSLOK-----RTLSDIENELKAL-----CYVAIRKENLPN 1616
QY 912 -ENFLRELKSEEM-----DVLGDKMEATEVEKLFQ-----RALPIR 948
Db 1617 LEKPIVDNASKNDVLNLCRSFSLVPLSTETEDYNMRKEHTKIILNGDPSIDFLKCEKY 1676
QY 949 KMAVATH-----CFTSPCHDIRFFSDGIGRIGITRKRHNGITFLKILPPTQSD 1001
Db 1677 QMLISKHYEKOEAENPGYEFLEKASALGY-----ELVSEVELDRMKQMDSPD 1729
QY 1002 LRTGGRSSRPLSAFRSGFSKGFIDFVPLPSKNEKLTAPLLKLVLGVACAFILVPSA 1061
Db 1730 IDYMQEKAARN-----EMVLL--RNEEKE-----ALQKIEYPSLFLIERAA 1770
QY 1062 -----DAVDALKTACALLKGRIBELAKCIANPACAAVACLOTQNNRPDETEQIKG 1114
Db 1771 GMKILVDQIEXDET-----IRK-----NHPTREMELESCH 1802
QY 1115 DLFENSVDENFCASVRKKCVPRK--SDLGERPAPDPSVLVONFNISDNGKWIYITSGIN 1173
Db 1803 HL--NLVLLDQNEYSYTLRPLENRRVEDLINLSKINYTAIPTIYQDLIGKYE-----N 1855
QY 1174 PTDAFDQCLHEPHEGDKNKLGNISWRIKTLDSDGFTFSRAVQFVQDPNPQGVLY---- 1229

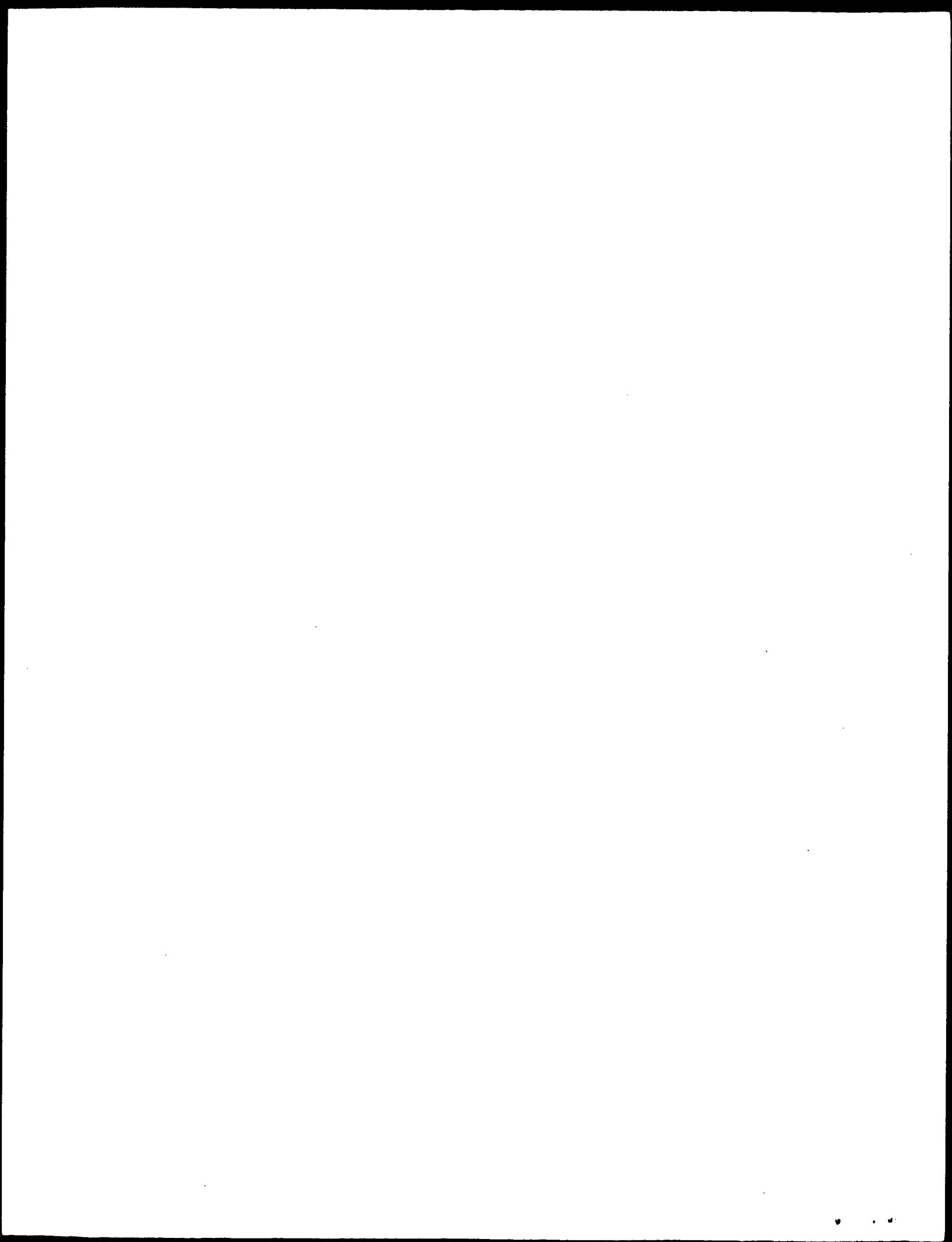
Db 1856 PNFDYLDKSLNK-----MDYVAISRQDYELMVAKYERQDLYLKIS 1896
QY 1230 -----NHDNEYLYHQDDWYILSSKIENKPEDYIFVYIRGRNDADWG 1270
Db 1897 SEKIDHIVVPLSEYNLMVTNVRNPSLSYLKKEKAVLNHILIKEDDYKNIL-----AVSE 1950
QY 1271 YGVAVVYTRSSVLPNSIPELEKAASKIGRDSFIFIRDNTCGPEPALVERIEKTVERGE 1330
Db 1951 HPTVHLSEKASLLANKVLVDKD-----DFATMSRSIE---KPTIDFLSTKALSMG- 1997
QY 1331 RIIVKEV-----EETEEVEKEVEKVGRTMTLQRLAEGFENELKQDEENFVRELKKEEM 1385
Db 1998 KILVNETHSKRNEKLLSEPDSEFLTMKAKEOGLIISKEYSELRDQIDRSLDLVREKA 2057
QY 1386 EFLDEIKMEASEVEKLFKALP 1407
Db 2058 AIFDSIIVENIEYQQLVNTTSP 2079
RESULT 15
S54174
DNA topoisomerase (EC 5.99.1.2) - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jun-2000
C:Accession: S54174
R:Tosh, K.T.; Kilbey, B.J.K.
submitted to the EMBL Data Library, January 1995
A:Description: Isolation and characterisation of the topoisomerase 1 gene from Plasmodi
A:Reference number: S54174
A:Accession: S54174
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <TOS>
A:Cross-references: EMBL:X83758; NID:g790481; PID:g790482
C:Superfamily: eukaryotic type I DNA topoisomerase
C:Keywords: DNA binding; DNA replication; isomerase

Query Match 1.98; Score 144.5; DB 2; Length 839;
Best Local Similarity 17.3%; Pred. No. 1.1;
Matches 191; Conservative 176; Mismatches 390; Indels 347; Gaps 46;

QY 352 LOKAASKVGRDFNFIITDNSCGPEPPLVERLEKTAEGEKKLLKEAVEI---EEVEKEV 409
Db 24 INKTONIG---NN--KSCNSRSKSKESIKKQNSLGIKNTKSLGKIKKEEKXQI 78
QY 410 EKVRDTEMTLFORLEGEKELQDEENFVRELKKEKEIINELQMEATEVEKLFGRALPI 469
Db 79 SKRSNELKKNLKEGKKYVEKKSRTV---KDEKLTNVIKKEQNNKK-----PK 128
QY 470 KRLMALAPHNSFNLANHETIKYVGVSKLPGHKRFSWGWDYFSGIVVAKICSSRRIPRYF 529
Db 129 KLLAKS---EENF---EPINRW--WEKIDQTDIQWNY----- 158
QY 530 RKSPPICGDLRGLQFLSHGKHNLSNPAHSINQNVKPGNSCKPEKPDVALMVWKEWQOFA 589
Db 159 -----LHRGL-IF-----SPPY-VQHHVP-----IPYKSIKTEL-----NA 188
QY 590 KTAIVAFILSVASKADAVALKTKTCLLKECRLELAKCISNPACAAVACLOTQNNRPD 649
Db 189 KSEELATVWCS-----AIGSDYCTKEKFIINFFKTFIN-----SLENDNIQK 231
QY 650 ETEQIKCGDLFENSVDENFCASVRKKCVPRKSDVGDGFPVDPSPVLVQKFDN----- 703
Db 232 ENETKLLKKGDISNFKFID-----FMPKIDHLLKLEKLNKTKEE 271
QY 704 KDFSGKWFITRGLNPTDAEDCQLHEPHEENKLVGNLSWRIRTPDGG---FFTRSAY 758
Db 272 KEKKKRWMEKELPYTALVDWIREKISSNAEPPLGFRGGEHPKQGLLKKRIFPDDVV 331
QY 759 OKFVQDPKYPGILYN-----HDNE--YLLYQDDWYILSKSVENSPEDYIFVYIKG 806

```
Db 332 INTSDAPVPRLYDNNCGNHWGDIYHDKVTVLWYKD-----SINDQIKYTFLS 381
QY 807 RNDWDGYGGSVLYTRSAVLPEIIPLOTAQAQKVGGRDENTFIKT-----DNTCGPEPPLV 862
Db 382 AOSKFKGYKDLMKYENAR-----KLASCYHKIREDYKNMKKNKIIDKQLGTAVYLI 433
QY 863 ERLEKKEVEGERTIIKEVEIEEVEKEVKVDKVTFLPSKLFEGFKELQORDEENFLRELSKE 922
Db 434 DFLALRV-GGEK-----DIIDEADV-----GCCSLRVEHISFAHDIPFK 472
QY 923 EMDVLDGLKMEATEVEKLGFRALPIRKLMVATHCFTS---PCHDRIRFSSDDGIGRLG 979
Db 473 SVD-----SKEQKTNDKVKNIPLPTNLESISEDYITLDFLGKDSIRYFN-----520
QY 980 ITRKRINGTFLKILPPIOSADLRTTGGRSSRPLSFRSGFSKGIFDIYVLPFSKNE-LKE 1038
Db 521 ---VKIDKQAYINII-----IFCKNKNRDEGVFDQITCSKLNLEYLKE 559
QY 1039 LTAPLLLLKLVGLACAFILIVPSADAYDAL--KTCACLLKGCRIELAKCIANPACAAVAC 1096
Db 560 IMPTLSAKVFRTYNASITLDQQLKRIKEVYGTYSLSYG-ETELHK-----SK 607
QY 1097 LOTCNRNPDETEQIRCGDLFENSVDNECAVSRK----KCVPRKSDLGEFPAPDPSV 1152
Db 608 KRKSSHLTSDTNILSDASDSTINDVNNYDENGINKKLSYATTVGKENDVDKNSP-IEV 666
QY 1153 LVQNFNISDPNGKWYITSGLNPTFDADCOLHEFTEGDNKLVGNISWRIKTLDSGFFTR 1212
Db 667 DVSNNIN-----ELINFYNNANREVAAILCNHQRSIPKQHDTTM 703
QY 1213 SAVQKFVODPNQPGVLYNHD-NEYLHYDDWYILSSKIENKPEDYIFVYVYGRNDWDGY 1271
Db 704 SKIKKQIE-----LYNEDIKEYKY-----LQHLKKNSDKKFIFV-----738
QY 1272 GGAVVYTRSSVLPNSIIPLEKAAKSGRDFSTFIRTDNTCGPEPALVERIEKTVEEGER 1331
Db 739 -----SKVSTLDGTLRP-----NKVKENMK 758
QY 1332 IIVKEVEIEEVEKEVEKVGRTMTLFORLAGFNLKODEENFVRELSKEEMEFDE- 1390
Db 759 -----EESCKKKL-----ITLIKVELLNNQMKVRRDDNKTALGTSKINTMDPR 802
QY 1391 -----IKMEASEVEKLGKALPIR 1409
Db 803 ITVAFCKKFEIPIEKVFNRLRK 826
```

Search completed: November 6, 2001, 05:03:40
Job time: 6764 sec




```
QY 337 IYTRSPITLPIESIIPIINLOKAASKVGRDNFNFTITDSCGPEPLVERLEKTAEGEKLK 396
Db 545 -----ISNEHSSLSOLSTIAEKEAAVATNB-----LSEKNSLOT 581
QY 397 EAVEIEEVEKEVKYRDTEM-----TLFORLLEGFKELOODENFVRELKKEKEILN 450
Db 582 LCNAFOEKLAKSVMLKENEQNFSSLDTSFKKLNESHQELNNHTQITKQL-KDTSSKLQ 640
QY 451 ELQMEATEVEK-----LFGRLPIRLKRLMALAPH-----SNFLANHETIYYVGS 495
Db 641 QLQERANFEOKESTLSDENNDRITKLLKLEESNKSILKKQEDVDSLEKNIQTLK----- 695
QY 496 KLPGRKFRSWSGWDYFGSIWAKICSSRIPIRYFRKSPRICGLDRLGOLF---SHGKH 552
Db 696 -----ED-----LRKSEALRFSK-----LEAKNLREVIDNLKGH 726
QY 553 NLSPA-----HSI-----NONVPKNGSGCKPKDVALVMVWEKGFQAKTAIVAIFILSVA 602
Db 727 ETLEAQNDLHSSLSDAKNTNAILSELTKSSEDV-----KRLTANVETLTQD 774
QY 603 SKA-----DAVDALKTCTCLLKECELELAKCISNACAAVACLOTNNRPDETECOIK 656
Db 775 SKAMKQFSLVNSYOSISNLYHELRRDHVN-----MQSQNTLLESESKLK 821
QY 657 --CGDLFEN--SYVDEF-----NECAVGRKKCVPRK-----SDVGDFPVPDP 694
Db 822 TDCENLTQNMWTLIDNVQKLMHKHVNQESKVELKNGKSLDLKLNKRLSSNLVAISDND 881
QY 695 SVLVQKFDW-KDFSGKWFITRGLNPTFDQDLQHEFTHEE-----NKLGNLSWRIR 746
Db 882 QILTQALSKNYSLEQSAQNLNGLSKLEAKQOLLHTENELHRLDKLTGKLGK----- 937
QY 747 TPDGFFTRSVAQKVFQDPKPYGIIYNHNDNEYLLODDWYILSSKVENSPEDYIFVYK 806
Db 938 -----IEESKSDL-----GKLTARQEE---ISNKEENMSQQAITSVKS 976
QY 807 RNDWDGCGSVLYTRSVAVLPESIPELOTAQKVGORFNTPIKTDNCGPPEPLVERLE 866
Db 977 KLDE-----TLKSSKL-BADIEHLKANKVSEVERNALLASN-----ERLM 1017
QY 867 KKEVEGERTIIEVEIEEVEKVRDKVTLFSLK-----FEGF-----KELORD 911
Db 1018 DDLKNNGE-----NTASLQTEIEKRAENDLQSKLSVVSSEYENLLLISSOTNKSLE-D 1071
QY 912 ENFLRELKSEMDVLGLUKMEATEVEKL---FGR-----ALPIRLMAVATHC 957
Db 1072 KTNQLYIEKNVQKLLDEKDQDNVLEELTSKYGLGEENAQIKDELLALRKK----- 1124
QY 958 FTSPCHDRIRFTSSD-----DGIGRL-----GITRKRINGTFLKILPPIQS--AD 1001
Db 1125 -SKQHDLCANFVDDLKESDALEQLTNEKNELIVSLQSNNSNEALVEERSDLANRLSD 1183
QY 1002 LRTTGRSRRLPSAFRSGFSKGFIDVPLPSPKNELKELTAPLLKLVLGVACAFVLPSPA 1061
Db 1184 MKKSLSDSNVISVIRD-----LVRVN 1206
QY 1062 DAVIDALKTACILKGCRIELAKCIANPACAAVACLOTNNRPDETECOIKGDLFENS 1121
Db 1207 DELDTLK-----KDKDSLSTQYSEVCQDRDDLDSLKGC----- 1240
QY 1122 VDEFNCAVS-RKKCVPRKSDILGEFPAPDPSVLVON---FNISDFNGKWIIT-SGLNPTFD 1177
Db 1241 ERSFNKYAVSLRELCTKSIDV-----PVSEILDNDNFVNAGNFSELSRLTVLSLENYLD 1295
QY 1178 AFDCQLHEFTEHCKNLGNISWRIKTLDGSGFFTRSVAQKVFQDPNQPGLYNHNDNEYL 1237
Db 1296 AFN-QVNFKKMELDN-----RLTTDAEF-----TKVVADE-----EKLOH 1330
QY 1238 YODDWKILSSKIENKPEDYIFVYVYGRNDWDGAGVAVYTRSSVLPNSIIPLEKAAS 1297
Db 1331 EHDDWLI-----QRG-----DUEKALKD 1348
```

```
QY 1298 IGRDFTSTRTDTCGPALVERIEKTEVEGRRIIVKEVEIEEVE----- 1345
Db 1349 SEKNF-----LRKEAEMTENIH-SLEGKEETKKEIAELSSRLDQLATNLKNO 1398
QY 1346 -----KEV---EKVGRTEMTLFORLAEGFNELQODENFVRELKKEMEPL--DEIKMEA 1395
Db 1399 LDHNLQETIRLKEDVLKESLIISLESLSNQKQKESLLD--AKNELEHMLDTSRKNS 1456
QY 1396 SEVEKL 1401
Db 1457 SLMEKI 1462

RESULT 2
RBP1_PLAVB STANDARD; PRT; 2869 AA.
ID RBP1_PLAVB STANDARD; PRT; 2869 AA.
AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE RETICULOCYTE BINDING PROTEIN 1 PRECURSOR.
GN RBP1.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=92315338; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
merozoites."
RL Cell 69:1213-1226(1992).
CC -I- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -I- SUBUNIT: HOMODIMER (POTENTIAL).
CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M88097; AAA29743.1; -.
DR HSSP; P36956; IAM9.
KW Malaria; Receptor; Signal; Transmembrane.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205BECFF CRC64;

Query Match 2.1%; Score 155.5; DB 1; Length 2869;
Best Local Similarity 17.4%; Pred. NO. 0.86;
Matches 283; Conservative 242; Mismatches 573; Indels 529; Gaps 76;

QY 57 TSYKTSFSDSSCHCKDKSQICSDTSTFEET-----QRFDLKRGMTLLILEKQWRQFIQLA 110
Db 1150 TSNKNEILKSVKEVDKLNLEQNEQDYKVKVKNPENKQLEAIRGSKLKE----- 1200
QY 111 IVLVCTFVIVPRYDAVDALKTACILKGCRIELAKCIANPACAAVACLOTNNRPDETE 170
Db 1201 -----VINKHVSEMTQLESTANTLK-----SNAGKNEHDELELN---KTK 1239
QY 171 CQTKGDLFE--NSVDQFNCAVSRKKKVPKRSVDGEPVDPDRA---VVQNFNM-KDF 224
Db 1240 GOMR--DIYEKLKKAIBELKEGTVNELKDKANEKANKVE-PEPERNIIGHVLERITVEKDK 1296
```


SQ SEQUENCE 2469 AA; 280165 MW; D2B4B8DC08644CDB CRC64;

Query Match 2.1%; Score 155; DB 1; Length 2469;
 Best Local Similarity 17.3%; Pred. No. 0.75;
 Matches 230; Conservative 208; Mismatches 497; Indels 396; Gaps 58;

QY 271 PTRASVQTFVDDPDLGALYNHNDNEFLHYQDDWYILSSQIENKPPDDYIFVYYR----- 323
 DB 178 FSTSNVDAIEYLPSPNQVT--GSFLYFVPEKIGHSH-----YIMHYRIVNYEKL 228

QY 324 -GRNDWDGCGSVIY-----TRSPTEPESIIPLNLOKAASVGRDENFI 367
 DB 229 HGNIDLTQEGLLIEISPPNPKPTPTQPPKTPRTPKATPRKPTPKRESTI 288

QY 368 TTDNSCGPEPLPLERLEKTEAGEKLIIKEAV-----EIEEVEKEVEKVRDT--EMTL 419
 DB 289 PYDKS--KKPP--KIPKTSKKKKVLTQATLPQHKTIIEHLRELLPITETVEDNTL 343

QY 420 FQR-----LLGFPKELQ-----QDENFVRELKKEKEILNELQMEATEVEKL 462
 DB 344 FHPVERTPTGDSLLSGINSTTKREDDEDONVTSKLKEDDDWIDD----- 392

QY 463 FGRALPIRKLRMALAPHNSFLANHETIKYVVGSKLPCHKRFSWGWEDYFGSIVVA----- 517
 DB 393 ----IPIPEVLDTTHS----DOETI-YMIGDE--NIHDSYSDDDDIDTDLDSFIQLD 441

QY 518 -KICSSRRIPRYFRKSPRICGLDSRGLQFSGK-----HNLSPAHSINQNPVKG 567
 DB 442 NLITSLDNIPKN--TFPRI--IDKTSNQPIKEGKALHSIDRIKLNIVLHGLTSSSIS 497

QY 568 NSGCKFPKVDALMWKQWQAFATVAIVFILSVASKADAVDAKTCCTLLKCEKRELEAK 627
 DB 498 ISKCKSLQFVIL--WGE-----KLSIPTR-DLTKILKTEIITEIAETALTK 542

QY 628 CISNPACAAVACIQTNNRPDETECOIKCGDLFENSVDNEENCAVSRKCKVPRKSDVG 667
 DB 543 -LTN-----DIFRNNVITKLNKMLK-----SESVD 569

QY 688 DFP-----VPDPSVLQKFM-----KDFS-----GKWFITRGLNPTF 720
 DB 570 SYKHLALLNNIILKIOTIDTEIELKTLSTVFTSELGKDFSVVCTRKESETIMAAIKNLK 629

QY 721 DAFDCQHEFHTEENKLVGNLSRIPTPDGPFTRSAVOKF-----VQDPKYPG 769
 DB 630 EKISTRQELHTEENYQSVL-----TAMETFIQPIPLTRVIEIQPSKKAQ 675

QY 770 ILYNHNDNEYLLODDWYILSSKVENSPEDYIFVYKGRND---AWDYGSGSVLYTRSAVL 826
 DB 676 QL--HEKSKLVEQK---LTIDANNVLTOLLHTMKQDKTDISPAPD-----FTTVLKN 722

QY 827 PESIPELOTAQKVRD---FNTFTKTNTGCPPEPLVERLEKKVEEGERTI--KEVEE 882
 DB 723 IQSTLQLQTCVTDNLNIDKFKFISNTVOQLSYIGWEVAELSHSQWNPFPKADPIPLKILD 782

QY 883 IEIEVEKVRDEKVTFLSKLPEGKELQORDENFLRELSKEEMDVLQKMEATEVEKLF 942
 DB 783 IKKEIQVTTQK-----KNEETLSKILADVQTLLENAKQSDILSIPLQHYHTKACTING 837

QY 943 RALPIRKLMVAVHTFSPCHDIRFSSDDGIGRIGIT-----RKRINGTFLKLPLPIQ 998
 DB 838 E-----RENQFESLKNVTQKLSSTSEEFKLTLIDSTTLLENVQLQIQ 878

QY 999 S-ADLRTTGRRSRPLSAFSGSKG---IFDIVLPSPKSELKELTAPILL----- 1045
 DB 879 EISDILQSNQYIHQSETIQKQAFDKSNTIINNLIQINQOQKVTYVTPQMLIAVKRFLSA 938

QY 1046 -----KLVGVLAFLIPLVSADAVDAKTCALLKGCRIELAKCIANPACAAVAC 1096
 DB 939 KFRESNTICEIISITVSLGSLSSKSTTVEALKDALKSIDTLKEKL----- 983

QY 1097 LQTCNNRPDETE-----COIKCGDLFENSVDNEENCAVSRKCKVPRKSDLGEPAPD 1149
 DB 1097 LQTCNNRPDETE-----COIKCGDLFENSVDNEENCAVSRKCKVPRKSDLGEPAPD 1149

DB 984 --TAVDRPLKRELYNVIRKLOKQLKTLLEQQEFDNW-----KMEVDSF-VPT 1027

QY 1150 PSVLYQNF--NISDFNGKWYITSGNLNPTFDADFDCOLHEFHTEGDN-KLVGNISWRIKTLTD 1206
 DB 1028 PSRDVKTFTFIONAPSKAKOYAKKALKDQIQAMEIDVDVESVIEDNIKANGQKAWQ----- 1082

QY 1207 SGFFTRSAVQKFEVQPNQPGVLYNHNDNEYLHYQDDWYILSSKVENKPKEDYIF-----VY 1261
 DB 1083 -----KIQSAFQDLNFSILI-----PDDWLSL-AKEYTRPKSTLFTVIGPILL 1124

QY 1262 RGRNDWDGCGAVVYTRSSVLPNSII-----PELEKAAK 1296
 DB 1125 KEVEEVLSEVKNLKEAKLSLLPNPGVFTPPKFDWHIYVESNVNFHLKTLINLPKSVSTVAH 1184

QY 1297 SIGRDFSTFIRTDNT-CGPEPALVERIE-----KTVE-----EGEIIIVKEV 1337
 DB 1185 NIGHELSSLLSQALNSKTLPEAVVGTSLSEQHAAKFSCKMKTLEATWHHDHOVTRTKIDYI 1244

QY 1338 EEIEEVEVK-----EVEKVR-----TEMTLFORIAEGFNELKQDEENFVRELKSEMEFL 1388
 DB 1245 EDLRNDTRKKHIVAPQIQSPNRFSLPEDIQEIINSLPKLFRDLSLENSRLASQKNEFOML 1304

QY 1389 DEIKMEASEVE 1399
 DB 1305 -ENTVKAALQLQ 1314

RESULT 4
 MYSB_PIG STANDARD; PRT: 1935 AA.
 AC P79293;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE MYOSIN HEAVY CHAIN, CARDIAC MUSCLE BETA ISOFORM.
 GN MYH7
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID:9823;
 RN [1]
 RC STRAIN-DOMESTICA;
 RA KO Y.L.;
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: MUSCLE CONTRACTION.
 CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
 CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -!- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE
 MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.
 CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U75316; AAC37320.1; -.
 DR InterPro; IPR000048; -.

QY 256 VGNLTWRIKTLGGFTRSAVQ--TFVQ-----D 282
Db : : : : :
QY 486 QQFNNHMFVLEQEEYKKGIEWTDFDGMQLQACIDILIEKPMGIMSLIEECMPKATD 545
Db : : : : :
QY 283 PDLPGALY-NHNEFLHYODDNYILSSQIENKPDYIFVYGRND-----AWDGYGGSVI 337
Db : : : : :
QY 546 MTFKAKLYDNHLGKSNFQKPRNKGQEAH-----FSLIHYAGTVDYNLGW-----L 594
Db : : : : :
QY 338 YTRSPPLPSIIPNLQKAA-KSVGRDFNNFITDSCGPEPPVLEKTAEEGKLLIK 396
Db : : : : :
QY 595 QKNKDPLETVVGLYQKSLKLLSNLFANY-----AGADAP-VDKGKGAKKGSFOTV 647
Db : : : : :
QY 397 EAVEIEEEVEKEVEKVRDFTMTLFORLLEGFKELOQDDENFVRELSKEKE-----447
Db : : : : :
QY 648 SALH-RENKMLTNLRST-----HPHFVRCIIPNETKSPGVMDNPL 688
Db : : : : :
QY 448 ILNELQMEAT-EVEKLFGRALP-----IRKRLMALAPHNFIANHETIKYKVS 495
Db : : : : :
QY 689 VMHQLRCNGVLEGIRCTCRKGFNRLYGDFRQYRILNPAIPEGQFIDSRKGAELKLS 748
Db : : : : :
QY 496 KLPGRKFRSGWEDYFGSVIVAKICSSRRIPRYFRKSPRICGLDRLGLQLFSGKHNLS 555
Db : : : : :
QY 749 LDIDHNOYFEGHTKFFKAGILGLLEEMRDELSRIITRI--QAQSRGVLSEMEFKLLE 806
Db : : : : :
QY 556 PAHS-----INONVPKNGSGCKFKPKDVALMVWEKQGFATAIVAIFILSVAKADAV-----608
Db : : : : :
QY 807 RRDLSLIOWNIR-----AFMGVKNWPMWKLKFKIKPLKLSAETEKEMANNKE 854
Db : : : : :
QY 609 -----DALKTCTCLLKECRLELAKCISNACAAVACIOTCNRPDETECIQKGDLE 662
Db : : : : :
QY 855 EGRVKALEKSEARKELEEKWVSLQE-----KNDLQLOVQAEQDNLADAEERCDLIK 910
Db : : : : :
QY 663 NSV-----YDFNE-----CAVSRK---KCVPRKSDVDGDFPVPDPSVLVQKFD 702
Db : : : : :
QY 911 NKIQLEAKVKEMTERLEDEEMNAELTAKKRKLEDECSSELKRDIDDLTAKVEKER--968
Db : : : : :
QY 703 MKDFSKWITRGLNPTDFACQLHEFTENKLVNLSWRIPTPDGFFTRSAVQFV 762
Db : : : : :
QY 969 -----HATENK-VKNLTEEMAGLD-----ELIVKLT 993
Db : : : : :
QY 763 QDPKYPGILYNHDNEVLLYQDDWYILSSKVENSPEDYIFVYKGRNDAMDGYGGSVLVYTR 822
Db : : : : :
QY 994 KKKALQEAHQALDQAEEDKVNLTAKVKLEQV-----DDEGLSDQDK 1042
Db : : : : :
QY 823 SA-----VLPESIIPELQTAQKV-----RDF-----NTFIKTDNTCGP 857
Db : : : : :
QY 1043 KVRMDLERAKRLEGDLKLTQESIMDLNDKQQLDERLKKDFELNALNARIEDRQALGS 1102
Db : : : : :
QY 858 E-----PPLVERLEKKVEBERT-----IKEVEEIEEVEK-----889
Db : : : : :
QY 1103 QLOKKLQELQARIELEEL-EAERTARAKVEKLRSDLSRELEESERLEEAGGATSVQI 1161
Db : : : : :
QY 890 -----VRDKEVTLFSKLFEGFKELQORDEENFLREL 919
Db : : : : :
QY 1162 EMNKKREAEFOKMRDRLEBATLQHEATAAALKRKHADSVAELGEQIDNLRQVKQKLEK 1221
Db : : : : :
QY 920 SKEDMDVLDLQKMEATEVEKLFGRALPRKLMVATHCFTSPCHDRIRFFSDDGIGRLG 979
Db : : : : :
QY 1222 SEFKLEDD-----VTSNMEEIITAKANLEK-----CRT-----LEDQMEHRS-----RAE 1264
Db : : : : :
QY 980 ITRKINGTFLKILPPIQSADLRTTGGRSSRPL---SAFRSGFSKGFIDVIVLPFSKNEL 1036
Db : : : : :
QY 1265 ETQRSVNDL-----TRQAKLQTEGELSRLQDEKEALISQLTRG-----KLTYTQOL 1312
Db : : : : :
QY 1037 KLTAPLLKLVGLVACAFIIPVSADAVDALKTACILKGCRIELAKIANPACAAVAC 1096
Db : : : : :
QY 1313 EDLKRQLEEEVAKNALAHALQASRHDCLDREYQEEETEAKAEQLQVLSK--ANSEVAQ 1370
Db : : : : :
QY 1097 LQTCNNRPDETECIQKCGDLFNSVVDEFENCAVSKKCVPRKSDLGEPFAPDPSPVLVON 1156
Db : : : : :
QY 1371 WRT-----KYETDAIQRTTEELEAKKLAQRLQDAEE-----1402
Db : : : : :

QY 1157 FNISDFNGKWIITSGNLPTDADFDCQLHEFTEGDKNLGVNISWRIKTLDSGFFTRSAVQ 1216
Db : : : : :
QY 1403 -----AVEAVNAKCSSLEKTKHRLQNEIEDLWVD-----VERSNA 1438
Db : : : : :
QY 1217 KFVODPNQPGVLNHDNEYLHYODDNYILSSKTEKNPEDYIFVYGRNDAMDGYGAVV 1276
Db : : : : :
QY 1439 AAALDKKQ-----RNFDKILVEMKOKYESQSELESQKE-----ARSL 1477
Db : : : : :
QY 1277 YTRSSVLPSNLIPELE-----KAAKSIGRDFSTFIRTDNTCGPEPALVERIEKTVEEGE 1330
Db : : : : :
QY 1478 STEFLKUNAYELESLEHLETFKRENKMLQEIISDLTEQLGSTGKSHLEKIRKQLEAEK 1537
Db : : : : :
QY 1331 RIIVKEVEIEEVEKEVEKVRGRTMTLQ-----RLAEGFNEELKQDEENFVRELS--1381
Db : : : : :
QY 1538 LELQSALEEAASLEHLEGKILRAQLEFNQIKAEIERKLAEKDEEMEQAKRNLHVRVDSL 1597
Db : : : : :
QY 1382 -----KEEME-FLDEIKMEASEVEKLFCKA 1405
Db : : : : :
QY 1598 QTSIDAETRNEALRVYKMKMGDLNEMEIQLSHANMAAEA 1639
Db : : : : :
RESULT 8
MK21_YEAST STANDARD; PRT: 1025 AA.
ID MK21_YEAST AC Q12176;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RIBOSOME BIOGENESIS PROTEIN MAK21.
GN MAK21 OR YDR060W OR D4237 OR YD9609.14.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96381250; PubMed=8789263;
RA Brandt P., Ramlow S., Otto B., Bloecker H.;
RT "Nucleotide sequence analysis of a 32,500 bp region of the right arm
of Saccharomyces cerevisiae chromosome IV.";
RL Yeast 12:85-90(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S228C / AB972;
RA Hunt S., Bowman S., Harris D., Barrell B.G., Rajandream M.A.;
RT Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=99003241; PubMed=9786894;
RA Edskes H.K., Ohtake Y., Wickner R.B.;
RT "Mak21p of Saccharomyces cerevisiae, a homolog of human CAATP-binding
protein, is essential for 60 S ribosomal subunit biogenesis.";
RL J. Biol. Chem. 273:28912-28920(1998).
CC -1- FUNCTION: REQUIRED FOR 60S RIBOSOMAL SUBUNIT SYNTHESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE CBF / MAK21 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: X84162; CAA58976.1; --
CC EMBL: Z74356; CAA98878.1; --
CC EMBL: Z49209; CAA99089.1; --
CC SGD: S0002467; MAK21.
DR Ribosome biogenesis; Nuclear protein.
FT DOMAIN 74 77
FT POLY-ASP.
FT POLY-ASP.
FT DOMAIN 117 122
FT POLY-ASP.
FT POLY-GLU.
FT DOMAIN 979 986

SO	SEQUENCE	1025 AA; 116676 MW; PB80378727ED71D8 CRC64;	
	Query Match	1.9%; Score 139.5; DB 1; Length 1025;	
	Best Local Similarity	20.9%; Pred. No. 1.8;	
	Matches 159; Conservative	94; Mismatches 225; Indels 283; Gaps 41;	
QY	345	PESIINLQAKAASVGRDNFNFTTDSNCGPPEPLVERLEKTAEGEKLLIKEAVEI---	401
DB	406	PEQEFNLLRGLVGNKIG-DIDSKVSKASY-----LLLLEQAHPNMKSIVIDAIVDIALR	459
QY	402	-----EEVEKEVEKVRDTEMTLFORLLEGFKEKLEQDEEN--	436
DB	460	PNADYHTTYSVITLQTLKRSDESVANKLVK---TYFTLFEKFL-----INTDKONTN	511
QY	437	FVRELSK--EKEEILN-----ELQMEATEVE-----KLF-----GRALPIR	470
DB	512	GVYKSNKSYEEKRNKFKHGKSGVSKIEKTEVDELDEKNSKLFSAALLGINRAPFA	571
QY	471	KLMA-----LAPHSN-----LANHETIKYVVGSKLPKHGRFSWGWEYF	511
DB	572	QIPASVVEVHMETLFAITHSSNFTSIQALVLINQVTK-----AKLNSDRVYRTLYESLF	627
QY	512	GSIVAKICSSRRIPRYFKRSPRICGLDSRGLQFSGHKHNLSPAHSINQVPGNSGC	571
DB	628	D-----PRLVNS-----KQGIYL-----NLLY	645
QY	572	KPKDVALMVEKQGFATAIVAIFILSVASKADAVDAKTKCTCLLKECLELAKCISN	631
DB	646	KSUKQDALNV-ERVEAFVR-----ILQVCSHWLNVTGTGFFPLL-----IQIAKTV--	692
QY	632	PACAAVACIQTNNRPDETECIQKCGDLFENSVDDEFNECAVSRKKCVPRKSDVGDFPV	691
DB	693	-----POIKNLLT--NTPVDYE-----YESDAEEEGQDKIKRKYDGRKD-----	732
QY	692	PDSVLVQKFDKDFSGKWFITRGLNPTDFADFCQLHFPH-----TEENKLVGNLSWIRIT	747
DB	733	-----PKFANAESKSLWEINFIN-----HFHPTVKTYANAYVTGETEQIAK	774
QY	748	PDGGFFT-----RSAVQ-----KFVQDPKYPGLYN-----HDNEVLL	780
DB	775	PDGLFTLSHFLDFRVYRSAKQNTNARTGTSIQPLFSGSRVNSVLVKAASIMHD-OGPV	833
QY	781	YQDDWYLSKSVEN-SPEDYIFVY-----RGR-----NDANDGY	814
DB	834	NTEW--LTKVBEDIKPEKFFQYFTTKTAGDKKSKASNFSDSDDEMEINW---888	
QY	815	GGSVLTRSNVLPESIIPELOTAQKVGDRFTFKTNTCGPEPL--VERLEKKEVEG	872
DB	889	-SALVKSRRPVEDSDSDSELDAED-----DF-----SDSTSDPEKDAIDDEAKSEGS	938
QY	873	ERTIIKEVEIEIEVEKVRDKEVTLFSKLFEGKELQDEENFLRELSKERMDVLGKLM	932
DB	939	QES--QDEEGLDEIFYDGGQDNDKK-RSFAESSEDESSEEEKEEENKEVSAKRA	995
QY	933	EATEVEKLFGRALPIRKLMATAVTHCFTSPCHDRIRFFSSDD	973
DB	996	KKQKQKNML-KSLPV-----FASADD	1015
RESULT	9		
ID	MLP1_YEAST	STANDARD; PRT; 1875 AA.	
AC	Q02455;		
DT	01-OCT-1993 (Rel. 27, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DE	15-JUL-1998 (Rel. 36, Last annotation update)		
DE	MYOSIN-LIKE PROTEIN MLP1.		
GN	MLP1 OR YKR095W OR YKR415.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TaxID=4932;		

```
QY 506 GWEDYFGSIWAKICSSRRIPRYFRKSPRICGLDSRGQLFSGHKNLSPAHNSINQNPV 565
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 402 HLQNOIETIIVELEHKVPIINSFKERTDMLNENALALLLHSTNEKNKAKVELN--A 458
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 566 KNGSGCKFPKDVAMVWEKW-----GOFATAIVAIFILSVASKA 605
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 459 KNOKLVECDNLQTLTKRQDLRCQIQYLLITNSVNSDKGLRKEEQ--FIQIMQED 516
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 606 DAV---DALKTCTLLKRCR--LELARCISIPACAANVACLOTNNRPDETECOIKG- 658
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 517 DSTITESDSQKVVVTEKLVBFKNIIQOE-----KNAELKVVVRNLADKLESKEKSK 568
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 659 ---DLFENSVWDEFNECAVRKCKVPRKSDVGD-----FVPPDPSVLVQKFD 702
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 569 QSLKTESVNEAKEAITLKS---EKMDLESRIEELQKELEELKTSVPNEDASYSNVT 625
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 703 MKDFSKWFTITGLNPTDFAPCQLEHFEHTENKLVGNLSWRIPDPDGGFTTRSAQKVF 762
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 626 IKQ-----LTKRDLQESQVQDLQTRISQIT-----RESTEN-----MSLLNKEI 665
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 763 QPKYPGILYNHDNEY-----LLYQDDWILS-----SKVENSPEYIFVYVKG 806
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 666 QD-----LYDSKSDISIKLGEKSSRIIAEERFKLLSNTLDTLRAENDQLRKRDYLN 719
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 807 -----RNDADWGYG-----SVLYTRSAVLPESIIPELOTAQKVGROFNTFIKTD- 852
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 720 TILKQDSKTHETLNEYVSKSLSVETELLNKEE-----QKLRVHLKKNLQKEL 770
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 853 NTCGPE-----PPLVERLEKKVVEGERTIIKEVEEIEEVEKVRKQVTLFSKL 901
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 771 NKLSPEKSLRIMVTQLOTLQKREDELLLETRKSCQKQKIDEDALSSELK-KET---SQK 826
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 902 FEFGKELQDE-----ENFURELSKEMDVLGDKLMEATEVEKLFGRALPIRLMVA 954
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 827 DHHIKOLEEDNNSNIEWYQNKTEALKKDYESVITSVDSKQTDIEKLYKVKSLKEIE-- 884
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 955 THCFTSPCHDRIRFP-----SDDGIGRIGITKRKRINFTFLKILPPIQSADLRITG 1006
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 885 -----EDKIRLHTYNYMDETINDDSL-RKELEKSKINLTDAYSQIK--EYKDYET- 932
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1007 GRSSRPLSAFRSGFKGIFDVLPSKKNELKELTAPLLLLKVLGVACAFVLPSADAVDA 1066
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 933 --TSQSLOQTNSKLDSEFKDFT-----NQIKMLT-----DEKTS 964
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1067 LKTCACLLKGRIELAKTIANPACAANVACLOTNNRPDETECOIKCGDLFEN-----SV 1121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 965 LEDKISLLKEQMFNL-----NNELDLQKGMEXEKADFKKRISILQNNKNEVEAV 1014
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1122 VDFNECAVRKCKVPRKSDLGEPAPDPVSVLVQNFNISDFNGKWYITSGNLNPTFDAFC 1181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1015 KSEY-----ESKLSKIONDL-----DQOITYANTAQNNYEQELQKHADVSKTISELRE 1062
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1182 QLHEFHTEGDKLVGNLSWRIKTLDGFPTRSAVOKFVQDPNQPVLVHNDNXYLHYODD 1241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1063 QLHTY-----KQCVKTLN-----LSRQLENALKENEKS---WSSOKESLLEOLD 1104
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1242 WYLSKSKENKPE-----DYIFVYGRNDAWDGYG-----AVVYTRSVLPNSIIP 1289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1105 --LSNSRIEDLSQNKLLYDQIYTAADKEVNNSTNGPGLNNILTLRRERDILDTKVT 1162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1290 ELEKAASIGRDFSTF-----IRNTDTCGPEPALVERTEKTVEGERIIVKEVEIE 1341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1163 VAERDAKURQISLMDVLOQAPKLNLS-----RVEK-----ENHSSLIQOHDIM 1210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1342 EYEKEVEKVGRTMTFORLAEGFNELKQDENFVRELSKEMEFLDEIKMEASEVE 1399
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1211 EKL-NQLNLLRESNITL-----RNEL-ENNNKKKELQSE---LDKLNQVAPIE 1255
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
* RESULT 10
DMD_CHICK STANDARD; PRT; 3660 AA.
```

```
AC P11533;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DYSTROPHIN.
GN DMD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098331; PubMed=3062582;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "Nucleotide sequence of chicken dystrophin cDNA.";
RL Nucleic Acids Res. 16:11815-11815(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Muscle;
MEDLINE=89210800; PubMed=3072195;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "The chicken dystrophin cDNA: striking conservation of the C-terminal
coding and 3' untranslated regions between man and chicken.";
RL EMBO J. 7:4157-4162(1988).
CC -1- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE
PLASMA MEMBRANE.
CC -1- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
ABP-120, ASP-180, OR BETA-FODRIN).
CC -1- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
EMBL: X13369; CAA31746.1; -.
PIR: S02041; S02041.
DR HSPP: O01082; IAA2.
DR InterPro: IPR000433; -.
DR InterPro: IPR001202; -.
DR InterPro: IPR001589; -.
DR InterPro: IPR001715; -.
DR InterPro: IPR002017; -.
DR InterPro: IPR002349; -.
DR Pfam: PF00307; CH; 2.
DR Pfam: PF00397; WW; 1.
DR Pfam: PF00569; ZZ; 1.
DR Pfam: PF00435; spectrin; 22.
DR PRINTS: PR00403; WWDOMAIN.
DR PROSITE: PS00019; ACTININ_1; 1.
DR PROSITE: PS00020; ACTININ_2; 1.
DR PROSITE: PS00021; CH; 2.
DR PROSITE: PS01159; WW_DOMAIN_1; 1.
DR PROSITE: PS00020; WW_DOMAIN_2; 1.
DR Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
Repeat.
KW DOMAIN 1 244 ACTIN-BINDING.
FT DOMAIN 300 3000 26 SPECTRIN-LIKE REPEATS.
FT DOMAIN 3052 3085 WW.
FT DOMAIN 3086 3357 CYS-RICH.
FT VARIANT 1171 1171 MISSING.
FT VARIANT 1869 1869 Q -> H.
FT VARIANT 1885 1885 K -> R.
SQ SEQUENCE 3660 AA; 422874 MW; 85493DAF6D5B6D4A CRC64;
Query Match 1.8%; Score 135; DB 1; Length 3660;
Best Local Similarity 17.6%; Pred. No. 17;
```


Matches 245; Conservative 196; Mismatches 482; Indels 472; Gaps 64;

QY 71 KDKSOICSDTSFEIQRFDLKRGM--TLI-LEKWRQF-----IQAIYVLC 115
 Db 1923 KKKEDINAVNQAERLSKGAKAPEPTVLQSLKRRWDFESKFAOFRRLNVAIQTYLED 1982
 QY 116 TFIIVPRVDALDKTACCLLKECRIELAKCIANPSCAANVACLQTCNNRPDETECOIKC 175
 Db 1983 TTFVMTESMTVETVPSTYLAEL-LQLQALSEVERLNSPVLOAKD-----C 2030
 QY 176 GDLFENSVDQFNECAVSRKKCVPR-----KSDVGFEFVDPDRAVNVQFNKMF 224
 Db 2031 EDLLKQ-----EELCNKIKDCLGRLOQHIDIHSHKKTAPALQSATPRETANIQD-KLTQL 2083
 QY 225 SGKWTISGL-----NPTFD-----AFDCOLHEFH-----MENDKL-----VGNLT 260
 Db 2084 NSQEKYNKMYDRDQARDKSKERWLFHCMEKSFNEWLTFETEEKLSRAQTEAGDVGHVK 2143
 QY 261 WR--IKTLGGGFTTSVAVQTFVQDPDLFG-----287
 Db 2144 TKQFLQELQDGI---GRQTVVKTNTVTGERIEQSSAADANVLKEQLGNLNTRWQEICR 2200
 QY 288 -----ALYHNDNEFLHYQDDWTILSQIENKPDYIFVYGRNDAMWDGYGGSVIYTR 340
 Db 2201 QLVKRRKRIEENKILSEFQEDLNKLIWLLEETENVIAIPLPGNEDQLRDLCLGKVLKV 2260
 QY 341 SPTLPESITIPNLOKAASVGRDFNFTTNSCGPEPPVLERLEKTAEEGEKLLIKRAVE 400
 Db 2261 EELPHKGI--LKRNETGG-----TTLGSASLNERKHLESTLKEASRLRLKVSRD 2311
 QY 401 TEEVEKEVEKVRDTEMTLQRLLEGEKELQQDENEVRELSKEKEILNELOMEATEVE 460
 Db 2312 LPEK-QKEIE-----ILLKDFIELNQ-----INQLFWITPVK 2344
 QY 461 --KLFGR-----ALPTRKRLMALPHSNFLANHETIKYVYVSKLPLGKRFSGWEDYFG 512
 Db 2345 NOLEYNOVGPGAFDIKETEAA-----VOAKQPNVEE-----2377
 QY 513 SIIVAKICSSRRIPRYFRKSPRICGLDRLGLFSGHKNLSPAHISINQVPGKNSGCK 572
 Db 2378 --VLSKGC-----HLKKEP-----ATHPVKKKLEDLNAWK 2407
 QY 573 FPKDVALMWKMQQAKTAIVAIFILSVASKADAVDAKLTCTCLKECELELAKCISNP 632
 Db 2408 AINHLILQLKAEK-PTFEGEPALTPGVLT-SQGTVAVD-----QARVTKETTSFTP 2456
 QY 633 ACAANVACLQTCNNRPDETECOIKCGDLFENSVDDEFNEC-----AVSR--KKCVPRKS 684
 Db 2457 EMPSSV-----LLEVPAADFNKAWAELTDWLSRLDREIKAQRY 2495
 QY 685 DVGDFPVPDPSVLVQKPMKDFSGKWFITRGLNPTFDAPDCOLHEFHEEENKLVGNLSWR 744
 Db 2496 TVGLDLDINDMIKQANMQDLEOR-----RP-----QDELITAAQ-----NLKNK 2537
 QY 745 IRTPDGFFTRSAVQKF-VDDPKYPGLIYNHNE-VLLYODDYIILSSKVENSPEDYIFV 802
 Db 2538 TSNOEARTIITDREKIQSQWDDVHGVLQNRQOLHEMOKDSTQWLEAKQE---ABOVLE 2594
 QY 803 YKGRNDAMWDGYGGSVLYTRSAVLPESI-----IPELQTAQAKVGRDFN- 846
 Db 2595 QAKAKLESWK-----EISYTVKALKONKSELKQFSKEIROQMNIQVNDVALKPVPRDYSA 2650
 QY 847 -----TFIKTDNTCGPEPPLVERLEKKVERGERTI-----IKEYE 881
 Db 2651 DQTRKVELMTDNI-----NATWATINKRVISREAALESALLMQEFLYDLLEKFLAWLTEAE 2706
 QY 882 EIEEVEKVRDKEVTL-----FSKLEGEKELQD-----EEN---FLRELSKE 922
 Db 2707 TTANVLQDATHKTEKLEDPQWVRELKMQWDLQAEIDAHTDIFHNLNDENGQKILRSLEGS 2766
 QY 923 EMDV-----LDGLKMEATEVEKFLGRALPIRKLMVATHCFTSPCHDRIRFSSDDGIGR 977
 Db 2767 EDVALLQRLDNNFRWSELRK---KSLNTRSHLEAST-----DQWKR 2806

RESULT 11

MYSR_HUMAN

ID MYSR_HUMAN STANDARD; PRT; 1937 AA.

AC P13535; Q14910;

DT 01-JAN-1990 (Rel. 13, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE MYOSIN HEAVY CHAIN, PERINATAL SKELETAL MUSCLE.

GN MYH8.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Skeletal muscle;

RC MEDLINE=90323631; PubMed=2373371;

RA Karsch-Mizrachi I., Feghali R., Shows T.B. Jr., Leinwand L.A.;

RT "Generation of a full-length human perinatal myosin

heavy-chain-encoding cDNA.";

RL Gene 89:289-294(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Skeletal muscle;

RC MEDLINE=95324556; PubMed=7601129;

RA Jullian E.H., Kelly A.M., Pompidou A.J., Hoffman R., Schiaffino S.,

Stedman H.H., Rubinstein N.A.;

RT "Characterization of a human perinatal myosin heavy-chain

transcript.";

RL Eur. J. Biochem. 230:1001-1006(1995).

RN [3]

RP SEQUENCE OF 502-1937 FROM N.A.

RC TISSUE=Skeletal muscle;

RC MEDLINE=90235862; PubMed=1691980;

RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goedde H.W.,

Arnold H.H.;

RT "Identification of three developmentally controlled isoforms of human

myosin heavy chains.";

RL Eur. J. Biochem. 189:55-65(1990).

RN [4]

RP SEQUENCE OF 860-1937 FROM N.A.

RC MEDLINE=89234168; PubMed=2715179;

RA Feghali R., Leinwand L.A.;

RT "Molecular genetic characterization of a developmentally regulated

human perinatal myosin heavy chain.";

RL J. Cell Biol. 108:1791-1797(1989).

RN [5]

Db 637 NIHVPNSYKQENKQEPYLLVLKKEIDKLKVPMPKVESLNEEKNKNIKTEGOSDNSEPT 696
 QY 652 ECOI-----KCGDLFENSVDNEFCACVSRKKCVPRKSDVGDVFPVDPDSVLVQ-- 699
 Db 697 EGEITGQATTKPGQQAAGSALG-----DSVQAQAQEQKQAQP-----PVPVPVPEAKAQP 747
 QY 700 -----KFDKDFSGKWFITRGLNPTDFADFCOLHEFTEENKLVGLNLSWRI 745
 Db 748 TPAPVNNKTVNSKLDYLE-----KLYEFL-----WTSYIC 779
 QY 746 RTPDGGFFTRSAVQKFPVQDPKYPGILYNHD--NEYLLYQDDWYILSSKVEN-----SPED 798
 Db 780 H-----KY--ILVSHSTWNEKILQ--YKITKEESKLSLSCDPLD 815
 QY 799 YIF-----VYVYKGRDWDGGSV-----LYTRSAVL-----PESIIPELOTAQ 839
 Db 816 LLENIONNIPVMT-----SMFDSLNNLSOLFMEIYKEMVCNLYKLKDNKDKTNLLEAK 871
 QY 840 KVRGRDNTFIKTONTGPEP-PLVERLEKKVVEGERT-----IIRKEVEE----- 883
 Db 872 KV-----STSVKTLSSSSMQPLSLTPQDKPEVSANDDTSHSTNLSLKLLENLSLGNK 927
 QY 884 -----EEVEKVRDKVTLFSLKFGFKELQDEENFLRELKSEMDVLDGLK 931
 Db 928 NIYQELIGKSSSEFYKILKDSDFYNESFTNFVSKKADINLSNDESKR-----K 980
 QY 932 MEATEVEKLFGRALPIRKLMAVATHCFTSPCHDRIRFFSSDDGIGRLGTRKRINTFLL 991
 Db 981 LE-EDINKL-----KYLQLSFDLYNKKYKLEFLDKKTKVGYKMQIKKLT---LL 1029
 QY 992 KILPPIQSADLRTTGKSSRP-----LSAFRSGFSKGFIDIVPLSPKNEKELTAPLLKL 1048
 Db 1030 K-----EQLESKLSNPNKHLVQNFVFNK-----KKEAE----- 1061
 QY 1049 GVLACAPLIVPSADAVDAKTCACLLKGR--LELAKCIANPACAAVACLOTCNNRPDE 1106
 Db 1062 -----IAETENLTMLKHGKLVGYNYNGESSPLKTLSEISIGTNYASL 1110
 QY 1107 TECQIKGDLFENSVDNEFCACVSRKKCVPRKSDIGEPFAPDPSVLVQNFNISDFNGK 1166
 Db 1111 E--NFKVLSKLEGLKLDNLN---LEKKLSYLSGLHLLIA-ELKEVIRKNV----- 1157
 QY 1167 YITSLNPTDFADFCOLHEFTEENKLVGNISWRILKTLSDGFFTSVQKFPQD----- 1221
 Db 1158 ---TGNSPS-----ENNTDVNNALE---SYK-KFLPEGTDVATVVSSEGSDFLEOS 1201
 QY 1222 -PNQGVLYNHDNEYLHYQDDWYILSKYENKPEDYIFVYVYGRNDA-WDGYGGAVVYTR 1279
 Db 1202 QPKKPA-----STHVGAESNTITTSQNVDEVDVILVPIFGESEEDYDLGQVV--TG 1253
 QY 1280 SSVLPNSIPELEKAAGKIGRDFSTIRDTNCTGCPALVERIEKTVBERGRIIVKEVEE 1339
 Db 1254 EAVTSPVI-----DN-----ILSKIE---NEYEVLYLKPLAG 1282
 QY 1340 IEEVEKEVKGRTMTLQRLAEGFNLKQDEENFVRELSEEMEFLD 1389
 Db 1283 VYRSLKKOLE---NNVMTFNVNVDILNFRNKRNFKNVLESDLPIYKD 1329
 RESULT 13
 ID MSP1_PLAFW
 AC P04933; STANDARD; PRT: 1639 AA.
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
 DE (PMSA) (P195).
 GN MSP-1.
 OS Plasmodium falciparum (isolate Wellcome).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5848;
 RN [1]

RP SEQUENCE FROM N.A.
 EX MEDLINE=86014355; PubMed=2995820;
 RA Holder A.A., Lockyer M.J., Odink K.G., Sandhu J.S., Riveros-Moreno V.,
 RA Nicholls S.C., Hillman J., Davey L.B.V., Schwarz R.T.,
 RA Freeman R.R.;
 RT "Primary structure of the precursor to the three major surface
 RT antigens of Plasmodium falciparum merozoites.";
 RL Nature 317:270-273(1985).
 RN [2].
 RP REVISIONS.
 RA Holder A.A.;
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
 CC (POTENTIAL).
 CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42
 CC KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
 CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
 CC -----
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X02919; CAA26676.1; -.
 DR PIR: A24594; A24594.
 DR InterPro: IPR000561; -.
 DR Pfam: PF00008; BGF; 1.
 KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;
 KW Transmembrane; GPI-anchor.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 764 764 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 768 768 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 783 783 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 844 844 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 920 920 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 964 964 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1058 1058 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1165 1165 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1174 1174 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1445 1445 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1526 1526 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1639 AA; 187618 MW; 2C255B6616C87F6E CRC64;

Query Match 1.88; Score 134; DB 1; Length 1639;
 Best Local Similarity 18.0%; Pred. No. 6.7;
 Matches 246; Conservative 207; Mismatches 473; Indels 444; Gaps 70;

QY 184 VDQNECAVSRKKCVPRKSDVGEFVPVDRNAVQNFNMKDFSGKWIITSLNPTDFADFC 243
 Db 249 IENINELLEESKKTIDK-----NKNATKEEKKKLYQAQYDLS-----LYNK 290
 QY 244 QLHEFFHMENDKLGNLTWRKTLDGFFTRSAVQTF--VQDP-----DLPALYNHDNE 295
 Db 291 QLEEAH-----NLISVLEKRIIDLKKNENIKELLDKINEIKNPPANSNTPTNLLDKKK 346
 QY 296 FLHYQDDWYILSSQIENKPPDY-----IFVYVYGRNDWDGSGSVIYTRSPPLPESII 349
 Db 347 IEEHEKEIKEIAKTIKFNIDSLFTDPLEYLYLREKKNID--ISAKVETKESTEPNE-Y 403
 QY 350 PNLQAAKSVGRDNFNFTTDSNCGPEPPLVERLEKTAEECEKLLI-----KEAVEIEEE 404
 Db 404 PNGVYTPLSY-NDINNALNELNSFG---DLINPDFYKEPSKNIYTDNERKKFNEIKEK 459
 QY 405 VEKEVEKVRDTEMTLQRLAEGFNLKQDEENFVRELSEKEKEILNEL-----QMEAT 457
 Db 460 IKIEKKKI-----ESDKKSVEDRSKSLNDITKEVEKLLNELIYDSKFNNDILT 507


```

FT MOD_RES 128 128 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).
FT CONFLICT 966 966 D -> E (IN REF. 2).
FT CONFLICT 978 978 T -> TE (IN REF. 2).
FT CONFLICT 986 986 E -> Q (IN REF. 2).
FT CONFLICT 1008 1014 DLOAED -> ALEAKT (IN REF. 2).
FT CONFLICT 1057 1057 D -> Y (IN REF. 2).
FT CONFLICT 1060 1060 L -> V (IN REF. 2).
FT CONFLICT 1095 1095 D -> N (IN REF. 2).
FT CONFLICT 1217 1217 E -> D (IN REF. 2).
FT CONFLICT 1271 1271 D -> N (IN REF. 2).
FT CONFLICT 1327 1327 T -> A (IN REF. 2).
FT CONFLICT 1358 1358 C -> R (IN REF. 2).
FT CONFLICT 1504 1504 L -> V (IN REF. 2).
FT CONFLICT 1537 1537 M -> L (IN REF. 2).
FT CONFLICT 1556 1556 N -> K (IN REF. 2).
SQ SEQUENCE 1934 AA; 222928 MW; FDBAC58310B057D CRC64;

Query Match 1.8%; Score 133.5; DB 1; Length 1934;
Best Local Similarity 17.0%; Pred. No. 8.9; 525; Indels 493; Gaps 63;
Matches 251; Conservative 211; Mismatches 211;

QY 188 NECAVSRKCVPRKSDVGEPPVDPNR-----AVQNFNMKDFSGKWYITSGLNPT- 237
DB 390 NSADLLKGMCHPRVKVGNVYTKGVQVQVSYAIGALAKSVYEKMFN--WMVTR-INATL 446
QY 238 -----DFADFQOLIE--FHMENDKLVGNLTWRIKTLDDGGFTTSVAVQ 277
DB 447 ETQPROYFIGVLDIAGFEIFDFNSFQLCINFTEKQLQFFNNHMFVLEQEEYKKEGIE 506
QY 278 -TFVQ-----DPLDPLGALY-NHDNEFLPHYQDDW 303
DB 507 WTFIDFGMDQACIDILEKPMRIMSILEECMPKATDFTFKAKLDNHLGKSNNQKPR 566
QY 304 YILSSQIENKPDYIFYVYGRND---AWDGYGGSVYITRSTPLPESIPNIAKAA-KS 358
DB 567 NVKCKQEAH-----FSLVHAGTVYDNLGW-----LQNKDPLNETVGLYQKSSKL 615
QY 359 VGRDFNFITDNSCGPEPLVERLEKTAEGEKLIIKEAVEI--BEEVEKEVYRDE 416
DB 616 LSNLFANY-----AGADAP-VDKGKAKKKGSSF---QTVSVLHRENKMLNTRST- 664
QY 417 MTLFQRLLEGFKELQDDEENFVRLSEKKEE-----ILNELQMEAP-EVEKLFGR 466
DB 665 -----HPHFVRCIIIPNETKSPGVMDNPLVMHQLRNGVLEGIRICRG 707
QY 467 LP-----IRKRLMALPHSNFLANHETIKYIVGSKLPGHKRFSWGWEDYFGSIV 515
DB 708 FPNRILYGDQRQRYRIINPAAIPBQFIDSRKGAELKLLSLDIDHNQYKFGHTKVFFKAG 767
QY 516 VAKICSSRRIPRYFRKSPRICGLDSRGLQFLSHGKHNLSFPAHS---INONVPKNGSGCK 572
DB 768 LGLGLEMRDRLSRIITRI--QAQSGLLSRMEFKLLERRDSLLVIQWNI-RAPWGVK 824
QY 573 -FP-----KQVALMVWEKQWQAKTAIVAIIFILSVASKADAVDAKLT 613
DB 825 NWPWMKLYFKIKPLKLSAETEKEMATMK-BEFGF-----VKDALEK 864
QY 614 CTCLLKECRLELAKCINPACAAVACIOTCNRPDETECOIKCGDLFENS-----VDE 668
DB 865 SEARRKELEKMWVSLLOE-----KNDLOVOAOEQDNIADAEERCDQLIKNKIOLAKVKE 920
QY 669 FNE-----CAVSRK---KCVPRKSDVGDFFVPDPVSLVQKDFMKDFSGKWFIT 713
DB 921 MTERLEDEEMNAELTAKKRLEDECSLKLKIDDLLETLAKVEKD----- 967
QY 714 RGLNPTFADFQOLHEFHTEENKVLGNLSWIRTPDGGFTTSVAVQVQDPKYPGLIYN 773
DB 968 -----HATENK-VKNLTEEMAGLD-----ETIAKLTREKALQEAHQ 1003
QY 774 HDNEVLLYQDDWY--ILSSKVE-----NSPEDIYFVYVYKGRND---AWDGYGGSVLVTR 823

```

```

DB 1004 QALDDLOAEEDKVNTLTAKSVKLEBQVDDLEGSLEQEKVMDLERAKRKLGLDILQEQ 1063
QY 824 AVLPESTIPELOTAQKVG-----RDF-----NWFIKTDNTCGPE-----PPLVERL 865
DB 1064 SIM-----DLENDKOQLDEKLLKKKDFELNALNARIDEQALGSQLOKKLQELQARIEEL 1117
QY 866 EKKVEEGERT-----IKEVEEIEEVEK----- 889
DB 1118 EEELE-EAERTARAKVEKLRSLSLEEEISERLEBAGGATSVQIEMNKKREAEFQKMRD 1176
QY 890 -----VRDKEVTLFSKLFEGFKELORDENFRELSEKEMDVLDGLKMEAT 935
DB 1177 LEATLQHEATAAALRRKKHADSVAELEGEQIDNLRQVKLEKESEFKLEIDD-----VTS 1232
QY 936 EVEKLFGRALPIRKLMAVATHCTFSPCHDRIRFPSSDDGIGRLGITRKRINGTFLKILP 995
DB 1233 NMEQLIKAKANLEKM-----CRT--LEDQNHRS-----KAEETORSVNDL----- 1272
QY 996 PIQSAADLTGTGSRSRPL---SAFRSGFSKGIQFIDVPLPSKNELKELTAPLLKLGVLA 1052
DB 1273 TSQRAKLOTENGELSRLQDEKEALISQLTRG-----KLTYTQOLEDLKRQL----- 1318
QY 1053 CAFLIVPSADAVDALKTCCALLKGCRIELAKCIANPACAAVACIOTCNRPDETE--CQ 1110
DB 1319 -----EEVKAQNTLAHALQSARHD-----CDLLREQVEEETEAKAE 1355
QY 1111 IKC-----GDLFENSVVDEFNECAVSRKCKVPRKSDGLGEFPAPDPSPVLQNFN 1158
DB 1356 LQCVLSKANSEVAQWRTKYETDAIQTQTELEEAKKLAQRLQDAEE----- 1401
QY 1159 ISDFNGKWYITSGLNPTDFADQOLHEFHTEGDKNLKLVGNISWRIKTLDSGFTTSVAVOKF 1218
DB 1402 -----AVEAVNAKCSLEKTKHRLQNETEDLMVD-----VERSNAAAA 1439
QY 1219 VQDPNPQGVLYNHDEYLYHQDDWYLLSSKIENKPEYIFVYVYGRNDADWDGYGGVAVYT 1278
DB 1440 ALDKQO-----RNFQKILAEWKQKVEESESLEESQKE-----ARSLST 1478
QY 1279 RSVLPNSLIPELE-----KAAKSIGRDFSTFTRDNTCGPEPALVERIEKTVVEGERI 1332
DB 1479 ELFLKLNAYEESLEHLETFKRNKNLQEEISDLTBOLGSGTKSIHELEKIRKQLEAKME 1538
QY 1333 IKVEVEETEEVEKEVEKVGRTMTLQ-----RLAEGFENLKQDEENFVRELS----- 1381
DB 1539 LOSALEEAEASLEHEEGNILRAQLEFNQIKAEIERKLAEKDEEMEQAKRNHLRVVDSLOT 1598
QY 1382 -----KEEME-FLDEIKMEASEVEKLFKA 1405
DB 1599 SLDAETRSNEALRVKMKMEGDLNEMEIQLSHANRMAAEA 1638

RESULT 15
LEF_BACAN STANDARD; PRT; 809 AA.
AC P15917;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE LETHAL FACTOR PRECURSOR (EC 3.4.24.-) (LF).
GN LEF.
OS Bacillus anthracis.
OG Plasmid pXOI.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-49.
RX MEDLINE=90034185; PubMed=2509294;
RA Bragg T.S., Robertson D.L.;
RT "Nucleotide sequence and analysis of the lethal factor gene (lef)
from Bacillus anthracis.";
RL Gene 81:45-54(1989).

```

[2]
RP SEQUENCE FROM N.A.
RA Lowe J.;
RL Submitted (APR-1990) to the EMBL/GenBank/DBJ databases.
RN [3]
RP ZINC-BINDING.
RX MEDLINE=95154669; PubMed=7851740;
RA Kochi S.K., Schiavo G., Mock M., Montecucco C.;
RT "Zinc content of the Bacillus anthracis lethal factor.";
RL FEMS Microbiol. Lett. 124:343-348(1994).
CC -!- FUNCTION: ONE OF THE THREE PROTEINS COMPOSING THE ANTHRAX TOXIN,
CC AGENT WHICH INFECTS MANY MAMMALIAN SPECIES AND THAT MAY CAUSE
CC DEATH. LF IS THOUGHT TO BE A LETHAL FACTOR THAT, WHEN ASSOCIATED
CC WITH PA, CAUSES DEATH. LF IS NOT TOXIC BY ITSELF. PA IS THOUGHT TO
CC BIND TO RECEPTORS OF SENSITIVE EUKARYOTIC CELLS, THEREBY
CC FACILITATING THE INTERNALIZATION OF LF OR EF.
CC -!- SUBUNIT: SECRETED ANTHRAX TOXIN IS COMPOSED OF THREE DISTINCT
CC PROTEINS, A PROTECTIVE ANTIGEN (PA), A LETHAL FACTOR (LF) AND AN
CC EDEMA FACTOR (EF). NONE OF THESE IS TOXIC BY ITSELF.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: THE PA-BINDING REGION IS FOUND IN BOTH B.ANTHRACIS EF
CC AND LF.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M34 (ZINC
CC METALLOPROTEASE).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M29081; AAA79216.1; -;
DR EMBL; M30210; AAA22569.1; -;
DR PIR; JQ0032; JQ0032.
DR MEROPS; M34.001; -;
DR InterPro; IPR000130; -;
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolyase; Metalloprotease; Zinc; Toxin; Signal; Repeat; Plasmid.
FT SIGNAL 1 33
FT CHAIN 34 809 LETHAL FACTOR.
FT DOMAIN 34 293 PA-BINDING REGION (POTENTIAL).
FT DOMAIN 300 420 REPEATS.
FT METAL 719 719 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 720 720 POTENTIAL.
FT METAL 723 723 ZINC (CATALYTIC) (POTENTIAL).
SQ SEQUENCE 809 AA; 93786 MW; 8C16B4D727310AE CRC64;

Query Match 1.8%; Score 132; DB 1; Length 809;
Best Local Similarity 17.2%; Pred. No. 3.4;
Matches 139; Conservative 117; Mismatches 254; Indels 300; Gaps 31;

QY 733 BENKLVNLSNRIRTPGCGFFTRSAVKQFVQDPKYPGLYNHD---NEYLLYQDDWY--I 787
Db 105 EMYKAIGG---KIYVDGDIYTKHISLEALSSEDKKKIKDIYKQALLHEHYVYAKEGEPV 161
QY 788 LSSKVENSPEDYIFVYKGRNDAMDGYGGSVLYTRSAVLPESEIPELQAAQKVGRDFNT 847
Db 162 L---VIQSSDYVENTEALN-----VYIEIGKILSRDILSKINQPYQKFLDLVNT 209
QY 848 FKTDNTCG-----PEPLVERLEKKVEEGERTIIK----- 878
Db 210 IKNASDSGDQLFTNLQKHEPTDFSVFLEQNSNEVQEVFAKAFAYIIEPQHRDVLQLY 269
QY 879 -----EVEIEEEVEKVRDKEVTLFSKLFEGFKE----- 907
Db 270 APEAPNYMDKFNQOEINLSLEELDKDQRMLSRYEKWEKIKOHYQHWSDSLSEEGRLKKL 329
QY 908 ---LQRDENFRELKSEMDVLDGLKMEATEV---EKLFGRALPI-----RKL 950
Db 330 QIPIEPKDDIIHLSQSEKELLKRIQIDSSDFLSTBEKEFLKKLIQDIRDSLSEEEKEL 389

QY 951 MAVATHCFTSPCHDRIRFFSSDDGIGRLGIRKTRINGTFLLLKILPPIQSAD---LRTTG 1006
Db 390 L-----NRIOVDSSNP-----LSEK--EKEFLKKLKDIOQYDINQRLQDGTG 429
QY 1007 GRSRRLP-----SAFRSGFSKGIFDIIVPLPSKNELKELTAPLLKLVG 1049
Db 430 GLIDFSINLDVRKQYKRDIONIDALLHOSIGSTLYNKIYLYENMNINLTATL----- 483
QY 1050 VLACAFIIVPSADAVDALKTCACLLKGCRIELAKCIANPACAAVACLOTNNRPDETEC 1109
Db 484 -----GADLVDS----- 491
QY 1110 QIKCGDLFENSVDDE--FNECAVSRKKCYPRK---SDLGEFPAPDPVSLVQNFNISDFNG 1164
Db 492 -----DNTKINRGIFNEFKNEKYSISSNYMIYDINERPALDNERLKWRIQLSPDTR 543
QY 1165 KWYITSGLNPTFDADFQCLHEFHTEGDKLV--GNISWRIKTLDSGFFTRSAVQKRVQDP 1222
Db 544 AGYLENG-----KLIQRLNIGLEIKDVO----- 566
QY 1223 NQPGVLYNHNDNEYLHYQDDWYILSSKIENKPEDYIFVYRGRNDAMDGYGAVVYTR--S 1280
Db 567 ---IIKQSEKEYIRI--DAKVVVFKSKIDTKIQE---AQLNINOEWNKALGLPKYTKLIT 617
QY 1281 SVLPNSIIELEKAAKSGIRDFSTFRTDNTCGPEPALVERIEKTVEEGE--RIIVKEV-- 1337
Db 618 FNVHNRYSNIVESAYLILNWKNNIQSD-----LIKKVTNYLVDGNGRFVFTDITL 669
QY 1338 -----EELEREEVEKEVKVGRTMTLFLORLAEGFNEKQDEENFVRE----- 1379
Db 670 PNIAEQYTHQDEIYEQVHSKGLVYVPSRSILLHGPKGV--ELRNDSEGIHFEFGHVAVDY 728
QY 1380 ---LSKEEME-----FLDEIKMEASEV 1398
Db 729 AGYLLDKNQSDLVNTNSKKFIDIFKEGGSNL 758

Search completed: November 6, 2001, 05:15:32
Job time: 671 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2001, 05:13:20 ; Search time 119.14 Seconds
(without alignments)
1568.027 Million cell updates/sec

Title: US-09-075-375A-6
Perfect score: 7495
Sequence: 1 MALSHTVFLCKEALNYA.....MEASEVKLFGKALPIKVR 1412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_16:*
- 1: sp.archaea:*
 - 2: sp.bacteria:*
 - 3: sp.fungi:*
 - 4: sp.human:*
 - 5: sp.invertebrate:*
 - 6: sp.mammal:*
 - 7: sp.mhc:*
 - 8: sp.organelle:*
 - 9: sp.phase:*
 - 10: sp.plant:*
 - 11: sp.rodent:*
 - 12: sp.unclassified:*
 - 13: sp.vertebrate:*
 - 14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2557	34.1	478	10 Q40593	Q40593 nicotiana t
2	2508	33.5	473	10 Q40251	Q40251 lactuca sat
3	2430	32.4	462	10 Q39249	Q39249 arabidopsis
4	1604.5	21.4	472	10 Q9SM43	Q9SM43 spinacia ol
5	1300	17.3	244	10 Q9SJD9	Q9SJD9 arabidopsis
6	1219.5	16.3	289	10 Q9FUS5	Q9FUS5 oryza sativ
7	210	2.8	522	10 Q9S713	Q9S713 arabidopsis
8	194.5	2.6	2269	5 Q26223	Q26223 plasmodium
9	161.5	2.2	2771	5 Q26216	Q26216 plasmodium
10	156	2.1	1365	2 Q49525	Q49525 mycoplasma
11	156	2.1	1933	13 Q90337	Q90337 cyprinus ca
12	155.5	2.1	886	1 Q29230	Q29230 archaeglob
13	151	2.0	1055	10 Q9S722	Q9S722 arabidopsis
14	151	2.0	1939	5 Q25662	Q25662 plasmodium
15	150.5	2.0	1332	4 Q9HAW4	Q9HAW4 homo sapien
16	148.5	2.0	2473	11 Q9QZ84	Q9QZ84 mus musculus
17	147.5	2.0	1786	5 Q9U0P0	Q9U0P0 plasmodium
18	146.5	2.0	1387	5 Q9GZ76	Q9GZ76 plasmodium
19	146	1.9	1935	4 Q9H1D5	Q9H1D5 homo sapien

20	146	1.9	2748	3 Q03767	Q03767 saccharomyc
21	145.5	1.9	1116	5 Q9VNH4	Q9VNH4 drosophila
22	144.5	1.9	839	5 Q26024	Q26024 plasmodium
23	144	1.9	2166	2 Q51465	Q51465 borrelia bu
24	144	1.9	3899	4 Q9Y6Y2	Q9Y6Y2 homo sapien
25	143.5	1.9	1558	5 Q96275	Q96275 plasmodium
26	143	1.9	1978	5 Q9W0M1	Q9W0M1 drosophila
27	142.5	1.9	880	1 Q9UZC8	Q9UZC8 pyrococcus
28	142.5	1.9	1199	5 P91349	P91349 caenorhabdi
29	142.5	1.9	1819	2 Q9ZLV0	Q9ZLV0 helicobacte
30	142.5	1.9	5105	5 Q61201	Q61201 caenorhabdi
31	141.5	1.9	1935	6 Q9GKRL	Q9GKRL sus scrofa
32	141	1.9	1002	2 Q66583	Q66583 aquifex aeo
33	140.5	1.9	1302	2 Q49547	Q49547 mycoplasma
34	140.5	1.9	1930	13 Q9DGD5	Q9DGD5 pennahia ar
35	140	1.9	800	1 Q59066	Q59066 methanococc
36	140	1.9	3595	4 Q9UQH3	Q9UQH3 homo sapien
37	139	1.9	2712	10 Q9SB74	Q9SB74 arabidopsis
38	139	1.9	3911	4 Q99996	Q99996 homo sapien
39	138.5	1.8	1327	4 Q9Y2L2	Q9Y2L2 homo sapien
40	138	1.8	1057	10 Q9F1I7	Q9F1I7 arabidopsis
41	138	1.8	2867	5 Q9N2M3	Q9N2M3 plasmodium
42	137.5	1.8	3908	4 Q9UQQ4	Q9UQQ4 homo sapien
43	137	1.8	1088	4 Q60772	Q60772 homo sapien
44	137	1.8	1109	6 Q00756	Q00756 oryctolagus
45	137	1.8	1676	10 Q23332	Q23332 arabidopsis

ALIGNMENTS

RESULT 1

Q40593 ID Q40593 PRELIMINARY; PRT; 478 AA.

AC Q40593;

DT 01-NOV-1996 (TREMREL. 01, Created)

DT 01-NOV-1996 (TREMREL. 01, Last sequence update)

DT 01-MAY-2000 (TREMREL. 13, Last annotation update)

DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.

GN TVDEL.

OS Nicotiana tabacum (Common tobacco).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;

OC Solanales; Solanaceae; Nicotiana.

OX NCBI_TaxID=4097;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=XANTHI; TISSUE=LEAF;

RA Bugos R.C., Yamamoto H.Y.;

RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL; U34817; AAC50031.1; -

DR Mendel; 9222; Nicotiana glauca; 9222.

DR InterPro; IPR000366; -

DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.

KW Transit peptide.

FT CHAIN 134 POTENTIAL.

FT TRANSIT 135 478 POTENTIAL.

SQ SEQUENCE 478 AA; 54561 MW; 0967DF4547D7809D CRC64;

Query Match 34.1%; Score 2557; DB 10; Length 478;

Best Local Similarity 100.0%; Pred. No. 6.9e-138;

Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 MALAPHSNPLANHETIKYVVGSKLFGHKRFSWGVEDYFGSIIVAKICSSRRIPRYFRKSP 533

Db 1 MALAPHSNPLANHETIKYVVGSKLFGHKRFSWGVEDYFGSIIVAKICSSRRIPRYFRKSP 60

QY 534 RICCGLDRLGQLFSGKHNLSPAHSINONPKNSGCKFPKDVALMYWEKQFAKTAI 593

Db 61 RICCGLDRLGQLFSGKHNLSPAHSINONPKNSGCKFPKDVALMYWEKQFAKTAI 120

QY 594 VAIFILSVASKADAVDAKTCCTCLLKECRLELAKCISNPACAANVACIQTCCNRPDTEC 653

```
Db 121 VAIFILSVASKADAVDAKTKCTCLLKECRLELAKISIPACAAANVACLOTNNRPDETEC 180
QY 654 QIKCGDLFNSVDFNECAVSRKCVPRKSDVGPDPFVSVLVQKDFMDKDFSGKWFIT 713
Db 181 QIKCGDLFNSVDFNECAVSRKCVPRKSDVGPDPFVSVLVQKDFMDKDFSGKWFIT 240
QY 714 RGLNPTFDADFQCLHEFTEENKLVNLSWRIRTPDGGFFTRSAVQKVPQKPGILYN 773
Db 241 RGLNPTFDADFQCLHEFTEENKLVNLSWRIRTPDGGFFTRSAVQKVPQKPGILYN 300
QY 774 HNEVYLLYODDWYILSSKVENSPEDYIFVYKGRNDANDGYSVLYTRSAVLPESIIPE 833
Db 301 HNEVYLLYODDWYILSSKVENSPEDYIFVYKGRNDANDGYSVLYTRSAVLPESIIPE 360
QY 834 LQTAAGKVGDRFNTFKTDNTCGPEPLVERLEKKVEGERTIIKEVEIEEVEKVRDK 893
Db 361 LQTAAGKVGDRFNTFKTDNTCGPEPLVERLEKKVEGERTIIKEVEIEEVEKVRDK 420
QY 894 EYTLFSKLFEGEKELODEENFLRSLKEEMDVLDGLKMEATEVEKLFGRALPIRKL 950
Db 421 EYTLFSKLFEGEKELODEENFLRSLKEEMDVLDGLKMEATEVEKLFGRALPIRKL 477

RESULT 2
Q40251 ID Q40251 PRELIMINARY; PRT; 473 AA.
AC Q40251;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN VDEL.
OS Lactuca sativa (Garden lettuce).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
OC euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.
OX NCBI_TaxID=4236;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ROMAINE;
RX MEDLINE=96270536; PubMed=8692813;
RA Bugos R.C., Yamamoto H.Y.;
RT "Molecular cloning of violaxanthin de-epoxidase from romaine lettuce
and expression in Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:6320-6325(1996).
DR EMBL; U31462; AAC49373.1; -.
DR Mendel; 8691; Lacsaa; Vdel; 8691.
DR InterPro; IPR000566; -.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Transit peptide.
FT TRANSIT 1 125 POTENTIAL.
FT CHAIN 126 473 VIOLAXANTHIN DE-EPOXIDASE.
SQ SEQUENCE 473 AA; 54447 MW; 1B2522DC2C62699 CRC64;

Query Match 33.5%; Score 2508; DB 10; Length 473;
Best Local Similarity 100.0%; Pred. No. 4.2e-135;
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MALSLHTVLCFEEALNLYARSPCNERFHRSGQPTNIIMKIRNNGVFNLSFRFTSYK 60
Db 1 MALSLHTVLCFEEALNLYARSPCNERFHRSGQPTNIIMKIRNNGVFNLSFRFTSYK 60
QY 61 TSSFSDSSCHKDKSQICSDTSTFEETQRFDLKRGMTLILEKQWRFQIQLAIVLCVTFVIV 120
Db 61 TSSFSDSSCHKDKSQICSDTSTFEETQRFDLKRGMTLILEKQWRFQIQLAIVLCVTFVIV 120
QY 121 PRVDADVDAKTKACCLKECRIELAKIANPSCAAANVACLOTNNRPDETECOIKCGDLFE 180
Db 121 PRVDADVDAKTKACCLKECRIELAKIANPSCAAANVACLOTNNRPDETECOIKCGDLFE 180
QY 181 NSVVDQFNECAVSRKCVPRKSDVGPDPFVSVLVQKDFMDKDFSGKWFITSGLNPTFDA 240
```

```
Db 181 NSVVDQFNECAVSRKCVPRKSDVGPDPFVSVLVQKDFMDKDFSGKWFITSGLNPTFDA 240
QY 241 FDCQLHEFHMDKLVNLTWRITKTLDDGGFFTRSAVQTFVQDDPLPGALYNHDNEFLHQ 300
Db 241 FDCQLHEFHMDKLVNLTWRITKTLDDGGFFTRSAVQTFVQDDPLPGALYNHDNEFLHQ 300
QY 301 DDMYILSSQIENKPDYIFVYKGRNDANDGYSVLYTRSPITPESIIIPNLQKAASVG 360
Db 301 DDMYILSSQIENKPDYIFVYKGRNDANDGYSVLYTRSPITPESIIIPNLQKAASVG 360
QY 361 RDNFNFTTNSGCGPEPLVERLEKTAEGEKLIIKEAVEIEEVEKEVEKVRDTEMTLF 420
Db 361 RDNFNFTTNSGCGPEPLVERLEKTAEGEKLIIKEAVEIEEVEKEVEKVRDTEMTLF 420
QY 421 ORLEGGFKELQDDEENFVRELSKEEKEILNELQMEATEVEKLFGRALPIRKL 473
Db 421 ORLEGGFKELQDDEENFVRELSKEEKEILNELQMEATEVEKLFGRALPIRKL 473

RESULT 3
Q39249 ID Q39249 PRELIMINARY; PRT; 462 AA.
AC Q39249;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN AVDEL.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Bugos R.C., Yamamoto H.Y.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
Khan S., Kim C., Altati H., Bei B., Chin C., Chioi J., Choi E.,
Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
Theologis A., Ecker J.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U44133; AAC50032.1; -.
DR EMBL; AC003981; AAF99753.1; -.
DR Mendel; 6341; Arath; Vdel; 6341.
DR InterPro; IPR000566; -.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Signal.
FT SIGNAL 1 113 POTENTIAL.
FT CHAIN 114 462 VIOLAXANTHIN DE-EPOXIDASE.
SQ SEQUENCE 462 AA; 52017 MW; 58E37B2C12D426B CRC64;

Query Match 32.4%; Score 2430; DB 10; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.1e-130;
Matches 462; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 951 MAVATHCFTSPCHDRIRFFSSDDGIGRLGTRKINGTFLKILPPIQSDADLTGGRSS 1010
Db 1 MAVATHCFTSPCHDRIRFFSSDDGIGRLGTRKINGTFLKILPPIQSDADLTGGRSS 60
QY 1011 RPLSAFRSGSKGIFDVLPLPSKNELKELTAPLLKLVGLVACAFIVPSADAVDAKTC 1070
Db 61 RPLSAFRSGSKGIFDVLPLPSKNELKELTAPLLKLVGLVACAFIVPSADAVDAKTC 120
QY 1071 ACLKKGRIELAKIANPACAAANVACLOTNNRPDETECOIKCGDLFNSVVDNECAV 1130
```

Db 121 ACLLKGRGRIELAKCIANPACAAVACIQTNNRPDETECOIKGDLFENSVDENECV 180
QY 1131 SRKKCVPRKSDLGFEFPAPDSVLVQNFNISDFNGKWIITSGLNPTDFADFCQLHEFHEG 1190
Db 181 SRKKCVPRKSDLGFEFPAPDSVLVQNFNISDFNGKWIITSGLNPTDFADFCQLHEFHEG 240
QY 1191 DNKLGVNISWRITLDSGFFTSRAVQKFDQPNQPGVLYNHNDNEYLHYQDDWYILSSKIE 1250
Db 241 DNKLGVNISWRITLDSGFFTSRAVQKFDQPNQPGVLYNHNDNEYLHYQDDWYILSSKIE 300
QY 1251 NKPDYIFVYRGRNDWDGYYGAVVYTRSSVLPNSIPELEKAASIGRDFSTFIRTDN 1310
Db 301 NKPDYIFVYRGRNDWDGYYGAVVYTRSSVLPNSIPELEKAASIGRDFSTFIRTDN 360
QY 1311 TCGPEPALVERIEKTVEGGERIIVKEVEEIEEVEKEVEKVGRTMTLQRLAEGFNEK 1370
Db 361 TCGPEPALVERIEKTVEGGERIIVKEVEEIEEVEKEVEKVGRTMTLQRLAEGFNEK 420
QY 1371 QDEENFVRELKSEMEFLDEIKMEASEVEKLFKALPIRKVR 1412
Db 421 QDEENFVRELKSEMEFLDEIKMEASEVEKLFKALPIRKVR 462

RESULT 4
Q9SM43
ID Q9SM43 PRELIMINARY; PRT; 472 AA.
AC Q9SM43;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE VIOLAXANTHIN DE-EPOXIDASE PRECURSOR.
GN SVDEL.
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
OC Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF.
RA Emanuelsson A.K., Eskling M., Akerlund H.E.;
RT "Cloning and sequencing of Spinacia oleracea violaxanthin de-
epoxidase.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ250433; CAB59211.1;
KW Signal.
FT SIGNAL 1 124 POTENTIAL.
FT CHAIN 125 472 VIOLAXANTHIN DE-EPOXIDASE.
SQ SEQUENCE 472 AA; 53658 MW; 12B4C8C69AB61E12 CRC64;

Query Match 21.4%; Score 1604.5; DB 10; Length 472;
Best Local Similarity 66.4%; Pred. No. 1.le-83;
Matches 299; Conservative 63; Mismatches 69; Indels 19; Gaps 5;
QY 968 FFSSDDGIGRLGTRKINGTFLKILPPIQSADLRTTGGSSRPLSAFRS-----GFSK 1022
Db 37 FLFQDDARRNIRFNDKLSCTKFICASEKIQ-----HSKSPKGLISCGVEVNSK 87
QY 1023 GIFDIVPLPSKNEKELTAPLLKLVGLVACAFILVPSADAVDAKLTACLLKGRGRIELA 1082
Db 88 VVSNAY-IPKKNLLKLVK---VVEVTAIVACTFFVMSAAQAVDAKLTCTCLLKGRGRIELA 143
QY 1083 KCIANPACAAVACIQTNNRPDETECOIKGDLFENSVDENECVSRKKCVPRKSD 1142
Db 144 KCIANPACAAVACIQTNNRPDETECOIKGDLFANKVDFEENECVSRKKCVPRKSDV 203
QY 1143 GEFPAPDSVLVQNFNISDFNGKWIITSGLNPTDFADFCQLHEFHEGDKLVGNISWRI 1202
Db 204 GEFPAPDSVLVQNFNISDFNGKWIITSGLNPTDFADFCQLHEFHEGDKLVGNISWRI 262
QY 1203 KTLDSGFFTSRAVQKFDQPNQPGVLYNHNDNEYLHYQDDWYILSSKIEKPEDYIFVYR 1262

Db 263 KTPDGGFFTRTAVQKFAQPSQPGMLYNHNDNAYLHYQDDWYILSSKIEKPEDYIFVYR 322
QY 1263 GRNDWDGYYGAVVYTRSSVLPNSIPELEKAASIGRDFSTFIRTDNCGPEPALVERI 1322
Db 323 GRNDWDGYYGAVVYTRSATVPENIVPELNRAAQSVGDKFNKFIKTDNCGPEPALVERI 382
QY 1323 EKTVEGERIIVKEVEEIEEVEKEVEKVGRTMTLQRLAEGFNEKODEENFVRELK 1382
Db 383 EKTVEGERIIVKEVEEIEEVEKEVEKVGRTMTLQRLAEGFNEKODEENFVRELK 442
QY 1383 EMEFLDEIKMEASEVEKLFKALPIRKVR 1412
Db 443 EMEFLDEIKMEASEVEKLFKALPIRKVR 472

RESULT 5
Q9SJD9
ID Q9SJD9 PRELIMINARY; PRT; 244 AA.
AC Q9SJD9;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE T27G7.23 (FRAGMENT).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altati H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howling B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T27G7 from chromosome
I.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006932; AAF22898.1;
DR InterPro: IPR000566;
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
FT NON_TER 244 244
SQ SEQUENCE 244 AA; 26631 MW; 4CD8221F444CF04C CRC64;

Query Match 17.3%; Score 1300; DB 10; Length 244;
Best Local Similarity 100.0%; Pred. No. 9.8e-67;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 951 MAVATHCTSPCHDRIRFFSSDDGIGRLGTRKINGTFLKILPPIQSADLRTTGGSS 1010
Db 1 MAVATHCTSPCHDRIRFFSSDDGIGRLGTRKINGTFLKILPPIQSADLRTTGGSS 60
QY 1011 RPLSAFRSGSGKIFDIVPLPSKNEKELTAPLLKLVGLVACAFILVPSADAVDAKLT 1070
Db 61 RPLSAFRSGSGKIFDIVPLPSKNEKELTAPLLKLVGLVACAFILVPSADAVDAKLT 120
QY 1071 ACLKGRGRIELAKCIANPACAAVACIQTNNRPDETECOIKGDLFENSVDENECV 1130
Db 121 ACLKGRGRIELAKCIANPACAAVACIQTNNRPDETECOIKGDLFENSVDENECV 180
QY 1131 SRKKCVPRKSDLGFEFPAPDSVLVQNFNISDFNGKWIITSGLNPTDFADFCQLHEFHEG 1190
Db 181 SRKKCVPRKSDLGFEFPAPDSVLVQNFNISDFNGKWIITSGLNPTDFADFCQLHEFHEG 240
QY 1191 DNKL 1194
Db 241 DNKL 244

RESULT 6
Q9FUS5 PRELIMINARY; PRT; 289 AA.
AC Q9FUS5;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE VIOLAXANTHIN DE-POXIDASE (FRAGMENT).
OS Oryza sativa subsp. indica.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
OC Oryza.
OX NCBI_TaxID=39946;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ZHENHUI-249;
RA Lin R.-C., Xu C.-C., Li L.-B., Kuang T.-Y.;
RT "Molecular cloning and expression of rice violaxanthin de-epoxidase";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF288196; AAF97601.2; -
FT NON_TER 1
SQ SEQUENCE 289 AA; 33496 MW; B91C7BBD4CE1770D CRC64;

Query Match 16.3%; Score 1219.5; DB 10; Length 289;
Best Local Similarity 74.8%; Pred. No. 4.7e-62;
Matches 217; Conservative 43; Mismatches 29; Indels 1; Gaps 1;

QY 1123 DEFNECAVSKKCPKRSKDLGEPAPDPSPVLVNFNFGKWKYITSGLNPTDFADCO 1182
|||||
Db 1 DEFNECAVSKKCPKRSKDLGEPAPDPSPVLVNFNFGKWKYITSGLNPTDFADCO 60
|||||

QY 1183 LHEFHTGDNKLVGNISWRKILKSDGFFTSVAVQKVPQDPNPGVLYNHDNEYLHYQDDW 1242
|||||
Db 61 LHEFVEGD-KLIANLTWRITRTPDSGFFRTAIOREVPQPAQPAIILYNDNEFLHYQDDW 119
|||||

QY 1243 YILSSKIENKPEYIFVYGRNDANDGAGVYVTRSSVLNLSIPELEKAASIGRDF 1302
|||||
Db 120 YILSSKIENKEDYIFVYGRNDANDGAGVYVTRSSVLNLSIPELEKAASIGRDF 179
|||||

QY 1303 STFRDTNCTGPEPALVERIEKVEGERIIVKEVEIEEVEKEVEKVGRTMTLFRKL 1362
|||||
Db 180 STFRDTNCTGPEPALVERIEKVEGERIIVKEVEIEEVEKEVEKVEEVEVILFKRL 239
|||||

QY 1363 AEGFNELKQDENFVRLSEKMEFEDEIKMEASEVEKFLGKALPKRKYR 1412
:|||||
Db 240 TDGLMEVKQDLNMFQGLSKEEMELLQNMNEATEVEKFSRALPIRKL 289
|||||

RESULT 7
Q9SJ13 PRELIMINARY; PRT; 522 AA.
AC Q9SJ13;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE AT2G21860 PROTEIN.
GN AT2G21860.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Unayun L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,

RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana";
RL Nature 402:761-768(1999).
DR EMBL; AC007019; AAD20404.1; -
SQ SEQUENCE 522 AA; 59287 MW; 63EF3102F3E96C21 CRC64;

Query Match 2.8%; Score 210; DB 10; Length 522;
Best Local Similarity 21.1%; Pred. No. 0.0003;
Matches 122; Conservative 77; Mismatches 186; Indels 194; Gaps 30;

QY 375 PEPLVERLEKTAPEGKLLIKAEVEIEEVEKEVEKVRDTEMTLQRLLEGKFKLQDE 434
|||||
Db 68 PEPPV-KLIALVGKGEVSPK-STSWEEVLMHTARKKWVD-----EGEMLVFDD 116
|||||

QY 435 ENFVRELSEKEKILNELQMEATEVEKLFGRALPIRKLALAPHSNFLANHETIKYVG 494
|||||
Db 117 E-----IVSSNDORAMN-LTQELNQTDLVVVA-----VNNSESV----- 150
|||||

QY 495 SKLPGHRRFSGWEDYFGSIIVAKICSSRRIPFRKSPRICGLDSRGLQLFSGKHNL 554
|||||
Db 151 -----NWTQTSKNVKNMIC-----FESSNLMNRLGG----- 178
|||||

QY 555 SPAHSINQVPGNSGCKFKPKDVALMV-----WEKVG-----OFAKTAIIVAIFILSV--- 601
|||||
Db 179 TDVGSVNRD-----KEVIEVKTVDWERNRNSDDIRFCLLVIIINAYIRPVPL 227
|||||

QY 602 ---ASKADAVDALKTCTCLKELELAKISNPACANVACIQTNNRPDETECIKCG 658
|||||
Db 228 QNLRSK-----GFSTLSOMVKNCGPQILNCLLDPCNKALQCLNQC-----PVDQVCYRCI 280
|||||

QY 659 DLFENSVDENECASVKKCPKRSKDVGPFPDPSPVLVQKFDKDFSGK----- 709
|||||
Db 281 ASYEGPYEAFSLCVLOKHNCLELDAKIPEKYVP-----MTSFRKELCHDTEAD 332
|||||

QY 710 -----WFTIRGLNPTDFADFCQLHEFH-----TEENKLVGN 740
|||||
Db 333 LFGVGLGLEWSVRVAGONPAYDQFCQYLFYRGKSKSFVPEVQVRLTEELV-- 390
|||||

QY 741 LSWRIPTDGGFFTRSAVKQVQDPKYP-----GILYNH--DNEYLLYQDDWYTLSSKVE 793
|||||
Db 391 ---WRRR-----RYSVKR-----GKIPATFRFSVLONGVVSNEF-----WTIVD--VS 428
|||||

QY 794 NSPEYIFVYKGRNDAMDGYGSLVYTRSAVLP-ESTIPELOTAQKVG-RDFTFIKT 851
|||||
Db 429 DDLWGLFHYHGAARVAGOSYTGAVLTPDGSYPAEKDKERLQSALEKCGIKEWELF-AV 487
|||||

QY 852 DNTCGPEPLVERLEKKEVEGER-----TIKEVEEIEE 885
|||||
Db 488 DNCSENPP-----GIQGSRLHSRISIEEPPDSEK 520
|||||

RESULT 8
Q26223 PRELIMINARY; PRT; 2269 AA.
AC Q26223;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE RHOPTRY PROTEIN.
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5862;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM;
RX MEDLINE=95021522; PubMed=7935623;
RA Keen J., Sinha K., Brown K., Holder A.;
RT "A gene coding for a high-molecular mass rhoptry protein of Plasmodium yoelii";
RL Mol. Biochem. Parasitol. 65:171-177(1994).
DR EMBL; L27838; AAA21304.1; -
SQ SEQUENCE 2269 AA; 265158 MW; F3D8CB103FB9A6E1 CRC64;


```
QY 119 IYPRVDVADALKTACALLKRECIELAKCIANPSCAAVACLQTCNNR-----PDTECOQK 174
Db 915 -LKKVD--DYIKV-----LNTNELITN-----CHNKQTLKDLKNQIK 951
QY 175 C-----GDLFENSVDQ-----FNECAVSRKCV-----PRKSDVGEF 207
Db 952 TIKETNSIDKIYTRDFENILTDKTELETKETGTLNHNESNNKELLTFYFDLKNLKG- 1010
QY 208 PVPDRNAVONFMKDFSGKWITSGLNPTDFADQCQLHEFHENDKLGNLTWRIKTLD 267
Db 1011 --NKENMLTKQNEKEKAV-----DIIKNVDINKIVSNIEITIT-- 1050
QY 268 GGFTRSAVQTFVQDPLGALYNHDN-----BFLHYQDDWYLS-----S 308
Db 1051 -----SYININEDTENEIGKSIELLNTK-----VLEKVKANVTNLN 1086
QY 309 QIENKPPDDYIFVYGRNDWDGCGSVIYTRSPPLPSIIPNLQAKAASVGRDNFNFI 368
Db 1087 EIKEKLKDYDFD-----GKEKNKYDENKIKNDITLQKIDKSITETLTKKN 1138
QY 369 TONSCGPEPLVERLEKTAEEGEKLLKEAVEIEEVEKEVGRDTEMTLFO-----RLLE 425
Db 1139 SENHIDEIKQIDKLKV--PNKTFWEDNEDPRELEKKIENIVEKI--DKKNYIKEIDKLLN 1195
QY 426 GKELQOODEN-----FVRELSKEKEI-----LNELOM 454
Db 1196 EISKIENDTSLKLNKINLSYKSGISLGNLFLOQIDEEKKAEHTIKAMEAYITDLDLNDIK 1255
QY 455 EATEVEKLFGRALPI-----RKLRLMALPHSNFLANHETIKYVVGSKLPGRKFS----- 504
Db 1256 KQOEIEKENNIMDKMDIHKEMKALINTSHDDYKIYHTTSKNH--BEKISDIRKNSLKIIQ 1314
QY 505 -WGVEDYFGSI-----YVAKICSRRIYFRKSPRICCGDLSRGLQLFSGHKHNLSP 556
Db 1315 DFSEESYINDIKKEKLVNLESONNNTDINOYLSKIENI-----YNILK 1358
QY 557 ASINONVPGNSGCKFPKDVVALVWKEWGQFATAIWAIFILSVASKADAVDAKLTCTC 616
Db 1359 LNKIKIIDD-----VKEYTDEIEKN-----NKKINAEIUSNSEKIITQ 1396
QY 617 L-----LKCRLELAKCISN---PACAAVACIQC-----NNRPDETECOIKCGDLFEN 663
Db 1397 LKENSILKECQSKIKSTIDNVSECIKNITNLKIYVNEKNN-----INTYFKN 1446
QY 664 SVVDENECASVSKKCVPRKSDVGFPPVDPVSVLQKDFMDKFSKWFITRGLNPTDFAF 723
Db 1447 A--BEYNQ-----NVSINFNNTIEMADTKSQYIL----- 1472
QY 724 DCQLHEFTEENKLVGNLSWRITPDGPFTRSAVQKVFQDPKPGIILYNHDNEVLLYQD 783
Db 1473 -----NKKNGTNTDYNKE-----LKEHKKSNVYKD 1502
QY 784 DWYILSKSVENSPEDYIFVYKGRNDWDGCGSVLYTRSAVLPESIIPELQTAQAQVGR 843
Db 1503 E-----AGKNTQE-----IKKNLEFKYEQEV-----TVLLNKYYAVELK----- 1538
QY 844 DFNFTIKTONTGCPPEPLVERLEKKEVEGERTIIEVEIEEVEKEVGRDKEVYLFSLKFE 903
Db 1539 --NKFDKTN-----YSEQIIEIKDAHNTFTSOADSEKKMEIKNEQIRI----- 1583
QY 904 GKELQORDEENFLRELSKEEMVDLD--GLKWEATEVEKLFGRALPIRKLMAVATHCF--TS 960
Db 1584 -----EDEVAKNNKSKAILDIQLSVPEPKI-----KFLKIDLTKSDCLKETK 1629
QY 961 PCHDRITRFSDDGIGRLGITRKRINGTFLKILPIQASDLRTTGGRRSRPLSAFRSGF 1020
Db 1630 DIETKISNISIDTQETKL--IENKNILNT--LEKLLSLKNO----- 1667
QY 1021 SKGIFDIVPLPSKNEKELTAPL-----LLKVLGVILACAPLIYPSADA 1063
Db 1668 KNKIED-----QKDELDEVNSKIKNIESNVHOKNKEIGIVEKINETAKA-----NKDQ 1717
QY 1064 VDALKTCALLKGRTELAKCIANPACAAVACLQTCNNRPD--ETECQ-----IKCGDL 1116
```

```
Db 1718 IESTQKLI-----IPTIKNLISPEFKANDLEGIDTNKLGKYNTEMNNIYEEFIKSYDL 1770
QY 1117 -----FENSVDDENECAVSRKCKVPRKSDLGEPAPDPSPVLVONF-- 1157
Db 1771 ITHYLETVSKPEITYGOIKNKRITAQNELLTNKYNKAKSYLDDIEANEFDRIVTHFKN 1830
QY 1158 NISDFNGKW---YITGLNPTFDACQLHEFTEGDKLVGNISWRIKTLDSGFFTRSA 1214
Db 1831 KLVNDVKFTNEY--SKVNGKFPDNISNINNVKSKTDENLLNLILNQTKEM-----YANIV 1884
QY 1215 VQKFPQDPNPGVLYNHDNEYLHYQDDWYILSSKIENKPEDEYIFVYGRNDWDGCGGA 1274
Db 1885 SKIYY-----SYKYEAEINIFINIPKLAN-----SUNIOIKSSSGI 1919
QY 1275 VVYTRSSVLPNSIIPLEKAAKSGIGHDFSTFIRTDNCGPEPALVERIEKIVBEGERI-- 1332
Db 1920 DLPKNINI---AILPYLDSOKK---DTLTFI-----PSP-----EKTSETYTKISD 1959
QY 1333 -----IVKEVEIEEVEKEVKEVGRTEMTLQRLAEGFNELKQDEENFVRELKSEEM 1385
Db 1960 SYNTLLDILKRSQLOKKEQOALNLI--FENRLHDKVOATNELK---DTLSDLKNKKE 2013
QY 1386 EFIDEIKM---EASEVEKL 1401
Db 2014 QILNKVKLLHLKSNELNKL 2032
RESULT 10
Q49525 PRELIMINARY; PRT; 1365 AA.
AC Q49525;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE LMP1.
GN LMP1.
OS Mycoplasma hominis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OC NCBI_TaxID=2098;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MH56;
RA MEDLINE=95369882; PubMed=7543881;
RX Jensen U.T., Ladefoged S., Birkelund S., Christiansen G.;
RT "Selection of Mycoplasma hominis PG21 deletion mutants by cultivation
RL in the presence of monoclonal antibody 552.";
DR EMBL: U21962; AAA81013.1; -.
SQ SEQUENCE 1365 AA; 154982 MW; 847A04992410867F CRC64;
Query Match 2.1%; Score 156; DB 2; Length 1365;
Best Local Similarity 16.8%; Pred. No. 1.3;
Matches 209; Conservative 187; Mismatches 420; Indels 430; Gaps 52;
QY 272 TRSAVQTFVQDPLGALYNHDNEFLHYQDDWYILSSQIENKPPDYIFVYGRNDWDG 331
Db 282 TRNOIQEFINT-----NKNNP--NYSE-----LISQLTSKRD-----SKNSVTD 319
QY 332 YGGSVIVYTRSPPLPESII--PNLQKA-----AKSVGRDNFNITDSCGPEPLVERLEK 385
Db 320 SNKSDIESANTELKQALAKANAKVQADNLAKSIKEQLNNSVSNANI-----LSA 369
QY 386 TAEGEKLLIKEAVEIEEVEKEVKEVGRDTEMTLQRLLEGF-----KELOODENFVREL 441
Db 370 KLTDKNTIOQAKTELEKEVQKADQAIKSNNTASQMSAKSSLDKVAEITKKLETFF--- 425
QY 442 SKEKEIILNELOMEATEVEKLFGRALPIRKLRLMALPHSNFLANHETIKYVVGSKLPGRH 501
Db 426 NKDEAKFNLKQTRNOIQEF----- 446
```



```

Db 642 EEIR 645
RESULT 13
Q9S722 PRELIMINARY; PRT; 1055 AA.
AC Q9S722;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE SMC-LIKE PROTEIN.
GN MIM.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=99380167; PubMed=10449416;
RA Mengiste T., Revenkova E., Bechtold N., Paszkowski J.;
RT "An SMC-like protein is required for efficient homologous
recombination in arabidopsis.";
RL EMBO J. 18:4505-4512(1999).
DR EMBL; AF120933; AAD54770.1;
DR EMBL; AF120932; AAD54769.1;
DR InterPro; IPR003439;
SQ SEQUENCE 1055 AA; 121349 MW; E9F0C0427FB602B4 CRC64;

Query Match 2.0%; Score 151; DB 10; Length 1055;
Best Local Similarity 19.3%; Pred. No. 1.8;
Matches 182; Conservative 119; Mismatches 306; Indels 334; Gaps 43;

Qy 128 ALKTCACLLKRECRYE-----LAKCIANPSCAANVACLOTCNNRPDTECOIKCGDLFEN 181
Db 57 AILTALCIAFCRARGTORAATLKDFTKTGSYAVVOVEMKNSGEDAFKSEIYGGVIIIE 116

Qy 182 SVDOFNECAVSR-----KKCVPRKSDVGEFPVPDPRNAVQNFNM-----KDFSGK 227
Db 117 RRTESATATVLDKLGKYSNKRDLRE-----LVEHFNIDVENPCVMSQDKAGS 168

Qy 228 WYLTSGLNPTFDPCQLHEFHENDKLGVNLWIRKTLDDGGFFTRSAVOTFFQDDPLPG 287
Db 169 SYI-----LECK-----GN-----SSSFLRNLLQ-----QVNDLLQ 194

Qy 288 ALYNDHNEFLHYODWYILSSQIEN--KDDYIFVYVRGRNDAMDGYGGSVIVTRSPPLP 345
Db 195 SIYELTKATAVD-----ELENTIKPIEKEISELRGK-----IKNWEQV 234

Qy 346 ESIIPNLQKAAK-----SVGRDFNNFITDNSCGPEPPLVERLEKTAEEGKLLIKE 397
Db 235 EEAQRLQQLKKLAWSWYDVGRQ-----LQEQTEKIVKLKERIPTCQ 278

Qy 398 AVETEEVEEVEKVRDT-----EMTLFORLEGFKE-----LQOD 433
Db 279 A-KIDWELGK-VESLROTLTKKAAQVACLMDDESTAMKREIESHQSAKTAVREKIALQEE 336

Qy 434 -----EENFVRELSKEKEILNELQMEATEVEKLFGRALP 468
Db 337 FNHKCNVQKIKDRVRLERQVGVDINQTMKNTQAEQSEIEELKYLREVEK----- 389

Qy 469 IRKRLMALAPHSN-----FLANHTIKYVGSKLPGHGRFSWGWDY 510
Db 390 VETLSRLKEBENFLEKAFEGRRKMEHIEDMKNHOKRQRFITSNINDLKKHQTKVTA 449

Qy 511 FGSIVWAKICSS-RRIPRYPRKSPRICGLDRLQGLFSGHKNLSPAHSINQVPGNS 569
Db 450 FGGDRVINLQAIERNHRRFKPP-----IGTGS----- 479

Qy 570 GCKFPKPDVALMWKQGFAKTAIVAIFILSVASKADAVDAKTKTCTLLKRECRLELAKCI 629

```

```

Db 480 -----HVTLVNGKWAASSVEQALGT--LLNAFIYTDHKOSL----- 513
Qy 630 SNPACAANVACLOTCNNRPDTECOIKCGDLFENSVDDEFNECAVS-RKKCVPRKSDVGD 688
Db 514 -----TLRGCAANEANYRNLI-----IYDFSRPLNPRHMYVPOTEHPTI 554
Qy 689 FVPDPSPVLVQKFDKMGDFSGKWFITRGLNPTF-----DAFDCOLHEFHTEENKLV--- 738
Db 555 FSVISD-----NPTFLNVLVQSGVERQVLAENYEEGKAVAFG 593
Qy 739 ---GNLSWRIRTPDG-GFFTRSAVQKFQD-PKYPGILYNHNDNEYLXYDDWYILSSKVE 793
Db 594 KRLSNLK-EVYITLDGKMFGRGPVQTTPLSRPSRLCASPDQI---KDLEIASKEQ 649
Qy 794 NSPEDYIFYVYGRNDAMDGYGGSVLYTRSAVLPSIPELOTAQAKVGRDPTFIKTDN 853
Db 650 NE-----INQMRKRREAENLEELKVRQLKRRSAEKVLTTKLEMHDLK-----N 699
Qy 854 TCGPEPPLVERL-EKKVEGERTIIKEVEIEEE---VEKVRD--KEYTL---FSKLPE 903
Db 700 TVAAE---IESLPSSSVNELQREIMKDLDEEIDEKEAFLEKQNCLEKAEELKANKULTALFE 756
Qy 904 GFKEQLQDE----ENFLRELSKEEMDVLGLKWEATEVEKL 940
Db 757 NMRESAKGEIDAFEAENELKKIEXD-----LQSAEAEKI 791

RESULT 14
Q25662 PRELIMINARY; PRT; 1939 AA.
AC Q25662;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)
DE REPEAT ORGANELLAR PROTEIN.
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=96V;
RA Werner E.B., Taylor W.R., Holder A.A.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U43145; AAC63403.1;
SQ SEQUENCE 1939 AA; 229001 MW; B36E462001C6F22F CRC64;

Query Match 2.0%; Score 151; DB 5; Length 1939;
Best Local Similarity 17.5%; Pred. No. 4.1;
Matches 189; Conservative 167; Mismatches 299; Indels 422; Gaps 50;

Qy 380 VERLEKTAEEGKLLIKEAVEIEEVEEVEKVRTEMTLFORLEGGFKELQOD---BEN 436
Db 95 VREYEVKIELEKEL---KLEKEKOINKEYEKELNEKSEFIKQOMELLKEKELINLKEN 151

Qy 437 FYRE-----LSKEEKEILNELQME---ATEVEKLFGRALPIRKLRMALAPHSNLANHE 487
Db 152 KINNKEIITLKREEK--LNDIESEYIEKNKEKEKL---NYEVTNIMKSL-----D 196

Qy 488 TIKYVGSKLPGHGRFSWGWDYFGSIVVAKICSSRRIPRYPRKSPRICGLDRLQGLF 547
Db 197 KLTCEVOEKDNLKIN-----KKVIEKENNLRELKEFKMEKNEIIESLDG----- 242

Qy 548 SHGKHNLSPAHSINQVPGNSGCKPDKVALMWKWKQGFAPKTAIVAIFILSVASKADA 607
Db 243 -----TIN-----DKNAYEKELEI-----SFEKRRM 264

Qy 608 VDALKTKTCTLLKRECRLELAKCISNPACANVACLOTCNNRPDTECOIKCGDLFENSVD 667
Db 265 IEMLD-----SKLIEKEENFAN-----KOAKLE 287

Qy 668 EFNECAVSRKCVPRKSDVGDFFPDPSPVLVQKFDKMGDFSGKWFITRGLNPTFDACQL 727

```

Db 288 KENIIEIKLDIESRE-----KDFKSK-----EKFASMENEL 321
QY 728 HEPHEENLVGNLS-WRIRTPDGGFFTRSAVOKFQDPKPGIILYNDHNEVLLYQDDWY 786
Db 322 NTLKSLDNACOMEVYKLEIKD---LQSOLVEKERE-----IFEIKNEY---DD-- 365
QY 787 ILSSKVENSPEDYIFVYKGRNDAMDGVGSLYTRSAVLPSIIPETAQAQKVRGDFN 846
Db 366 ----KINNMEKLSI-----ND--KGIDNTVLHSEEEKI-NKLKKEKETELNEHKKYN 413
QY 847 TFIKTONTGPEPPLVERLEKKVGERGERTIIEVEBIEEVEKVR-----DKEVT 896
Db 414 LEITE-----TIKNELNEKEEELKKAHTVEVTNLTKEIK 449
QY 897 LFSKLFEGFKELORD-----ENFLRELKSEMDVLDGLKMEATEVEKLFGRALPKRLMA 952
Db 450 LLEKKTEDAKGHKNELNELNQLSKLNEK-----DNIKNENTEL----- 490
QY 953 VATHCFTSPCHDRIRFFSDDGIGRITRKRINGFTFLIKILPPIQSDALRTTGGSSRP 1012
Db 491 -----NDKISSLNE-----VN-----ILNKDKQTLG----- 512
QY 1013 LSAFRSGFKGIFDVLPSKNEKELTAPLLKLVGLACAFILVPSADAVDAKTCAC 1072
Db 513 -----NDIKTLN-----TECOIKCGDLFNSVVD 527
QY 1073 LLKGRIELAKCIANPACAAVACLOTNNRPD-----TECOIKCGDLFNSVVD 1123
Db 528 -----EINTSDKNKKKEDLAMEWEGKC-----VVID 558
QY 1124 EF-----NECAVSRKCVPRK--SDLGEPAPDPVSVQNFISDNFGKWIYITGLNPTF 1176
Db 559 ETEKKYKEIFMELEKKEKENYADLND-----EISILRNSIVVKE--KEFI----- 603
QY 1177 DAFDCOLHEFHTEGDKLVGNISWRILKLDGFFTRSAVOK---FVQDPNQGVLYNHND 1233
Db 604 -----EMKEFY--ENKI-----NLFKNPFEKKNIYENLSRLKYDNEQ 642
QY 1234 EYLHYODDWILSSKLENKPEDYIFVYGRNDAMDGVGGAVVYTRSSVLPNSIPELEK 1293
Db 643 GLIKQIDELNIQKTEEK---YLQLY-----NDN-----MHMFRSICIKIDMPYSEN 687
QY 1294 AAKSIGRDFST-FI--RTDNTGPEPALVERIEKTVERGERIIVKEVEBIEEVEKEVEK 1350
Db 688 IKGSLVDVFTAYIKRDESSDANPDTHK--EMVAELEKRAHAIAVALEBKHEEIAK 745
QY 1351 VGTEMTLQRLAEGNE-----LKODEENFVRELKSEMEFDELKMEASEVEKLF 1403
Db 746 LGEGHKVYVLRIGEOHKEETIILEEKHKDVVTKLGEQHKENIILEEHEKDVVTKLG 802

RESULT 15
O9HAW4 PRELIMINARY; PRT: 1332 AA.
AC Q9HAW4
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HU-CLASPIN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21000493; PubMed=11090622;
RA Kumagai A., Dunphy W.G.;
RT "Claspin, a novel protein required for the activation of Chk1 during a
RL DNA replication checkpoint response in Xenopus egg extracts.";
RL Mol. Cell 6:839-849(2000).
DR EMBL: AF297866; AAG24515.1;
SQ SEQUENCE 1332 AA; 150175 MW; 4EA38DC16A3936C3 CRC64;

Query Match 2.0%; Score 150.5; DB 4; Length 1332;
Best Local Similarity 19.3%; Pred. No. 2.6;
Matches 255; Conservative 182; Mismatches 405; Indels 477; Gaps 67;
QY 307 SSQIENKPDYIFVYGRN-----DAWDGYSVYIYRSTPTLPESIPNLOKA 355
Db 83 SABEENKEN-----LYAGKNTKIRIVKTVDSDSEMSKLYQEN--LEAQVKPCLELS 135
QY 356 AKS-VGDRFNNFTTNSCGPEPPLVERLEKTAEBEKKLLIKEAVEIEEVEKEVEKVRD 414
Db 136 LQSGNSTDF-----TTDRKSKKH--IHDKEGTA---GKAVKSKRLEKE-ERKMEKIQ 185
QY 415 -----TEMTLFORLEGFKELO-ODEE-----NFVR 439
Db 186 LKKETKNOEDDVEQPNDSGCLLVKDLFETGLENNSPLEDESLESIRAANKVK 245
QY 440 ELSKEKEILNELQM--EATEVEKLFGRALPIRKLMA--LAPHSNLANHETIKYVGS 495
Db 246 KHKKEPSLESGVHSEFGSELSK--GTT---RKERKAARLSKEALKQLHSETQRLIRES 300
QY 496 --KLPGHKRESWGEDYFGSIVVAKICSSRPRIYPRKSPRICGLDSRGL-----QLFS 548
Db 301 ALNLPYH-----MPENKTIHDFKRPRTCHGNAMALLKSSKYQS 342
QY 549 HGKHNLSPAHSINQNVKNGSGCKFKPDVALMVWKGQFAKTAIVAFILSVASKADAV 608
Db 343 HKKEI IDTANTTENSDDHSGSE-----QITGAENEVETNA--LPVSKETQI 389
QY 609 DALKTCTLLKCRLELAKEISNPACAAVACLOTNNRPDETECOIKCGDLFNSVVD 668
Db 390 -----ITGSDESCRLVK-----NEELEIOEK----- 412
QY 669 FNECAVSRKCVPRKSDVGPVDPVSVLQKFDKMDKDFSGKWEITRGLNPTDAFCQLH 728
Db 413 -----OKQSDIRSP--GDSSVLOQESNF-----LG 436
QY 729 EFHTEENKLVNLSWRITPDG--GFTRSAVQRFVDP----- 765
Db 437 NNHSEECQVGLVAFEPHAGEGEPONPETDEKVEEPEQONKSSAVGPEKVRFTLDR 496
QY 766 -KYPGILYNHNEVLLYQDDWYILSSKVENSPEDYIFVYKGRNDAMDGVGSLVYT--- 821
Db 497 LKQGVDSVKPRIGADESFVILEPETNRELEALKQRFWKHANPAKPRAGTGVNVNI 556
QY 822 -----RSVLPESIIP-----ELQTAQKVG-----RDF----- 845
Db 557 VKDMGTGKEELKADVVVPTLAPKLDGASHTKPKGKQLVKAKLOEAMKLRFEERQKR 616
QY 846 NTFIKTDNTGPEPPLVERLEKKVEGERTIIEVEBIEEVEKEVEKVRDKEVTLFSKLFEGF 905
Db 617 QALFKLDNEDGSE-----EEEEEEEMTDESE-EDGEKVEK-EKEELEE-EGK 665
QY 906 KELORDEEN-----FLRELKSEMDVLDGLKME-----ATEVEKLFGRALPIRKLMAVA 954
Db 666 EEEEEEGNQETAFFL--LSSEETKDEKEMDKENNDGSSSEIGKAVGFLSVKPSLSSDS 723
QY 955 THCTSPCHDRIRF-----FSSDDGIGR-----LGITRKRINGTFL--KIL 994
Db 724 TLLLFKDSKMGYSPTKESETDENGSKQPSKLDDEDDCSLLTKESSHNSSEFELGSI 783
QY 995 PPIQSADLRITGGRSSRPLS-AFRSGFSKGIF--DIVPLPSKN--ELKELTAPLLKLKG 1049
Db 784 PSYQPCNRQTGRGTSFPPTAGGRSP--SPGLFRASLVSSASKSGKSLSPSLP----- 835
QY 1050 VLACAFILVPSADAVDAKTCACLLKCRLELAKEISNPACAAVACLOTNNRPDETE 1108
Db 836 -----IEDSQDLYNASPEPKTLFLG-----AGDFQFCLE-----DDTQ 868
QY 1109 CQTKGGLF-----ENSV---VDENECAVSR-----KKCV 1137
Db 869 SQLLDADGFLNVRHNRNQALKPRPLASMDENAMDNKDELIDLTGKFTSQAEKHL 928

```
Qy 1138 RKSD-----LGEFPADPPSVLVQNFNLSDFNGK-----WYITSG 1172
Db 929 RKSDKENMEELNLCSGKFTSQDASTPAS-----SELNKOKESSMGDPMEELALCSGS 984
Qy 1173 NPTFDADFQCLHEPHTGDNKLVGNISWRIKTLDSGFFTRSAVQKFVQDPNQPGLVYNHD 1232
Db 985 FPTDKEEDEEERF---GDFRLVSN-----DNEF---DSDEHSDSGNDLALDEHE 1030
Qy 1233 NEYLYQDDWYL---SSKIENKPEDYIFVYIRGRNDADWDGCGAVVYTRSSVLPNSIPE 1290
Db 1031 DD-----DEELLARSEKLEK-----ROMRLRKYLEDE 1057
Qy 1291 LEKAAKSIGRDFSTFIRTDNCTGPEPALVERIEKTVEEGERIIVKEVEEIEEVEKEVEK 1350
Db 1058 AEVSGSDVGSE-----DEYDGE-----IDEYEDVIDEVLPSDEELQSOIKK 1100
Qy 1351 V-----GRTEMTLFOR--LAEGFNLKQDENFVRELSKEEMEFLEIKMEASEVE 1399
Db 1101 IHMKTMLDDDKRQLRLYQERYLADG--DLHSDGPGRMKFRWKNIID-----DASQMD 1150
```

Search completed: November 6, 2001, 05:13:32
Job time: 586 sec

